PB89-234959

Methods for Evaluating the Attainment of Cleanup Standards. Volume 1. Soils and Solid Media

Westat Research, Inc., Rockville, MD

Prepared for:

Environmental Protection Agency, Washington, DC

Feb 89

U.S. Department of Commerce
National Technical Information Service

United States
Environmental Protection
Agency

Office of Policy, Planning,
and Evaluation
Washington, DC 20460

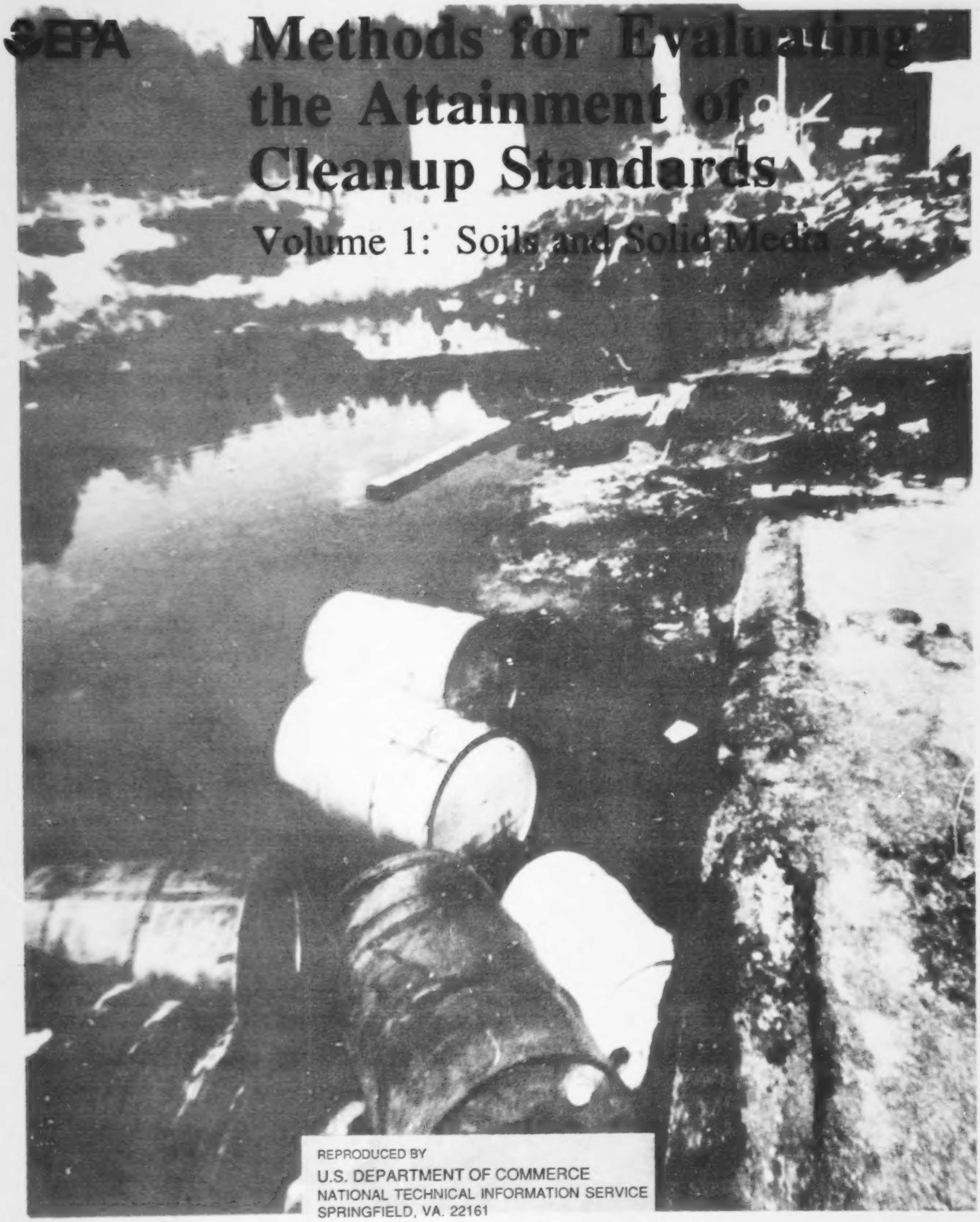
EPA 230/02-89-042
February 1989

PB89-234959

EPA

Methods for Evaluating the Attainment of Cleanup Standards

Volume 1: Soils and Solid Media



REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA. 22161

Methods for Evaluating the Attainment of Cleanup Standards
Volume 1: Soils and Solid Media

Statistical Policy Branch (PM-223)
Office of Policy, Planning, and Evaluation
U. S. Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460

DISCLAIMER

This report was prepared under contract to an agency of the United States Government. Neither the United States Government nor any of its employees, contractors, subcontractors, or their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use or the results of such use of any information, apparatus, product, model, formula, or process disclosed in this report, or represents that its use by such third party would not infringe on privately owned rights.

Publication of the data in this document does not signify that the contents necessarily reflect the joint or separate views and policies of each co-sponsoring agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TABLE OF CONTENTS

Page

Authors and Contributors xv
Executive Summary xvi

1. INTRODUCTION 1-1

- 1.1 General Scope and Features of the Guidance Document 1-1
  - 1.1.1 Purpose 1-1
  - 1.1.2 Intended Audience and Use 1-3
  - 1.1.3 Bibliography, Glossary, Boxes, Worksheets, Examples, and References to "Consult a Statistician" 1-6
- 1.2 A Categorization Scheme for Cleanup Standards 1-6
  - 1.2.1 Technology-Based Standards 1-7
  - 1.2.2 Background-Based Standards 1-7
  - 1.2.3 Risk-Based Standards 1-8
- 1.3 Use of this Guidance in Superfund Program Activities 1-8
  - 1.3.1 Emergency/Removal Action 1-8
  - 1.3.2 Remedial Response Activities 1-9
  - 1.3.3 Superfund Enforcement 1-9
- 1.4 Treatability Studies and Soils Treatment Technologies 1-9
  - 1.4.1 Laboratory/Bench-Scale Treatability Studies 1-10
  - 1.4.2 Field/Pilot-Scale Treatability Studies 1-10
  - 1.4.3 Soils Treatment by Chemical Modification 1-11
  - 1.4.4 Soils Treatment by In Situ Removal of Contaminants 1-11
  - 1.4.5 Soils Treatment by Incineration 1-13
  - 1.4.6 Soils Removal 1-13
  - 1.4.7 Soils Capping 1-14
- 1.5 Summary 1-14

2. INTRODUCTION TO STATISTICAL CONCEPTS AND DECISIONS 2-1

- 2.1 Hypothesis Formulation and Uncertainty 2-2
- 2.2 Power Curves as a Method of Expressing Uncertainty and Developing Sample Size Requirements 2-6
- 2.3 Attainment or Compliance Criteria 2-8
  - 2.3.1 Mean 2-9
  - 2.3.2 Proportions or Percentiles 2-9
- 2.4 Components of a Risk-Based Standard 2-11
- 2.5 Missing or Unusable Data, Detection Limits, Outliers 2-12
  - 2.5.1 Missing or Unusable Data 2-12
  - 2.5.2 Evaluation of Less-Than-Detection-Limit Data 2-15

TABLE OF CONTENTS

Page

2.5.3 Outliers....2-16
2.6 General Assumptions....2-17
2.7 A Note on Statistical Versus Field Sampling Terminology....2-17
2.8 Summary....2-18

3. SPECIFICATION OF ATTAINMENT OBJECTIVES....3-1

3.1 Specification of Sample Areas....3-1
3.2 Specification of Sample Collection and Handling Procedures....3-4
3.3 Specification of the Chemicals to be Tested....3-5
3.4 Specification of the Cleanup Standard....3-5
3.5 Selection of the Statistical Parameter to Compare with the Cleanup Standard....3-6
3.5.1 Selection Criteria for the Mean, Median, and Upper Percentile....3-6
3.5.2 Multiple Attainment Criteria....3-9
3.6 Decision Making With Uncertainty: The Chance of Concluding the Site Is Protective of Public Health and the Environment When It Is Actually Not Protective....3-10
3.7 Data Quality Objectives....3-11
3.8 Summary....3-12

4. DESIGN OF THE SAMPLING AND ANALYSIS PLAN....4-1

4.1 The Sampling Plan....4-1
4.1.1 Random Versus Systematic Sampling....4-2
4.1.2 Simple Versus Stratified Sampling....4-4
4.1.3 Sequential Sampling....4-6
4.2 The Analysis Plan....4-6
4.3 Summary....4-7

5. FIELD SAMPLING PROCEDURES....5-1

5.1 Determining the General Sampling Location....5-1
5.2 Selecting the Sample Coordinates for a Simple Random Sample....5-3
5.3 Selecting the Sample Coordinates for a Systematic Sample....5-5
5.3.1 An Alternative Method for Locating the Random Start Position for a Systematic Sample....5-10
5.4 Extension to Stratified Sampling....5-13
5.5 Field Procedures for Determining the Exact Sampling Location....5-13
5.6 Subsampling and Sampling Across Depth....5-14
5.6.1 Depth Discrete Sampling....5-15
5.6.2 Compositing Across Depth....5-15
5.6.3 Random Sampling Across Depth....5-17

TABLE OF CONTENTS

Page

5.7 Quality Assurance/Quality Control (QA/QC) in Handling the Sample During and After Collection 5-18
5.8 Summary 5-18

6. DETERMINING WHETHER THE MEAN CONCENTRATION OF THE SITE IS STATISTICALLY LESS THAN A CLEANUP STANDARD 6-1

6.1 Notation Used in This Chapter 6-1
6.2 Calculating the Mean, Variance, and Standard Deviation 6-2
6.3 Methods for Random Samples 6-4
  6.3.1 Estimating the Variability of the Chemical Concentration Measurements 6-4
    6.3.1.1 Use of Data from a Prior Study to Estimate σ 6-5
    6.3.1.2 Obtain Data to Estimate σ After a Remedial Action Pilot 6-5
    6.3.1.3 An Alternative Approximation for $\hat{\sigma}$ 6-6
  6.3.2 Formulae for Determining Sample Size 6-7
  6.3.3 Calculating the Mean, Standard Deviation, and Confidence Intervals 6-10
  6.3.4 Inference: Deciding Whether the Site Meets Cleanup Standards 6-11
6.4 Methods for Stratified Random Samples 6-12
  6.4.1 Sample Size Determination 6-13
  6.4.2 Calculation of the Mean and Confidence Intervals 6-15
  6.4.3 Inference: Deciding Whether the Site Meets Cleanup Standards 6-18
6.5 Methods for Systematic Samples 6-20
  6.5.1 Estimating Sample Size 6-20
  6.5.2 Concerns Associated with Estimating the Mean, Estimating the Variance, and Making Inference from a Systematic Sample 6-21
    6.5.2.1 Treating a Systematic Sample as a Random Sample 6-22
    6.5.2.2 Treating the Systematic Sample as a Stratified Sample 6-22
    6.5.2.3 Linearization and Estimates from Differences Between Adjacent Observations of a Systematic Sample 6-25
6.6 Using Composite Samples When Testing the Mean 6-26
6.7 Summary 6-27

7. DETERMINING WHETHER A PROPORTION OR PERCENTILE OF THE SITE IS LESS THAN A CLEANUP STANDARD 7-1

7.1 Notation Used in This Chapter 7-2

TABLE OF CONTENTS

Page

7.2 Steps to Correct for Laboratory Error.......7-3
7.3 Methods for Simple Random Samples.......7-5
  7.3.1 Sample Size Determination.......7-5
  7.3.2 Understanding Sample Size Requirements.......7-6
  7.3.3 Estimating the Proportion Contaminated and the Associated Standard Error.......7-7
  7.3.4 Inference: Deciding Whether a Specified Proportion of the Site is Less than a Cleanup Standard Using a Large Sample Normal Approximation.......7-8
  7.3.5 Deciding Whether a Specified Proportion of the Site is Less than the Cleanup Standard Using an Exact Test.......7-9
7.4 A Simple Exceedance Rule Method for Determining Whether a Site Attains the Cleanup Standard.......7-11
7.5 Methods for Stratified Samples.......7-12
  7.5.1 Sample Size Determination.......7-13
  7.5.2 Calculation of Basic Statistics.......7-16
  7.5.3 Inference: Deciding Whether the Site Meets Cleanup Standards.......7-19
7.6 Testing Percentiles from a Normal or Lognormal Population Using Tolerance Intervals.......7-20
  7.6.1 Sample Size Determination.......7-21
  7.6.2 Testing the Assumption of Normality.......7-23
  7.6.3 Inference: Deciding Whether the Site Meets Cleanup Standards Using Tolerance Limits.......7-24
7.7 Summary.......7-26

8. TESTING PERCENTILES AND PROPORTIONS USING SEQUENTIAL SAMPLING.......8-1

8.1 Notation Used in This Chapter.......8-2
8.2 Description of the Sequential Procedure.......8-3
8.3 Sampling Considerations in Sequential Testing.......8-4
8.4 Computational Aspects of Sequential Testing.......8-5
8.5 Inference: Deciding Whether the Site Meets Cleanup Standards.......8-7
8.6 Grouping Samples in Sequential Analysis.......8-8
8.7 Summary.......8-10

9. SEARCHING FOR HOT SPOTS.......9-1

9.1 Selected Literature that Describes Methods for Locating Hot Spots.......9-1
9.2 Sampling and Analysis Required to Search for Hot Spots.......9-1
  9.2.1 Basic Concepts.......9-1
  9.2.2 Choice of a Sampling Plan.......9-4
  9.2.3 Analysis Plan.......9-8
9.3 Summary.......9-8

TABLE OF CONTENTS

Page

10. THE USE OF GEOSTATISTICAL TECHNIQUES FOR EVALUATING THE ATTAINMENT OF CLEANUP STANDARDS 10-1

10.1 Background 10-2
10.1.1 What Is Geostatistics and How Does It Operate? 10-2
10.1.2 Introductory Geostatistical References 10-4
10.2 Soils Remediation Technology and the Use of Geostatistical Methods 10-4
10.2.1 Removal 10-5
10.2.2 Treatment Involving Homogenization 10-6
10.2.3 Flushing 10-7
10.3 Geostatistical Methods that Are Most Useful for Verifying the Completion of Cleanup 10-8
10.4 Implementation of Geostatistical Methods 10-9
10.5 Summary 10-12

BIBLIOGRAPHY BIB-1

APPENDIX A STATISTICAL TABLES A-1

APPENDIX B EXAMPLE WORKSHEETS B-1

APPENDIX C BLANK WORKSHEETS C-1

APPENDIX D GLOSSARY D-1

INDEX ....IND

LIST OF FIGURES

Page

Figure 1.1 Steps in Evaluating Whether a Site Has Attained the Cleanup Standard 1-2

Figure 2.1 A Statistical Perspective of the Sequence of Ground Water Monitoring Requirements Under RCRA 2-4

Figure 2.2 Hypothetical Power Curve 2-6

Figure 2.3 Hypothetical Power Curve Showing False Positive and False Negative Rates 2-7

Figure 2.4 Measures of Location: Mean, Median, 25th Percentile, 75th Percentile, and 95th Percentile for Three Distributions 2-10

Figure 2.5 Components of a Risk-Based Standard 2-13

Figure 3.1 Steps in Defining the Attainment Objectives 3-2

Figure 3.2 Geographic Areas and Subareas Within the Site 3-4

Figure 4.1 Illustration of Random, Systematic, and Stratified Sampling 4-3

Figure 5.1 Map of a Sample Area with a Coordinate System 5-2

Figure 5.2 Map of a Sample Area Showing Random Sampling Locations 5-6

Figure 5.3 Examples of a Square and a Triangular Grid for Systematic Sampling 5-6

Figure 5.4 Locating a Square Grid Systematic Sample 5-8

Figure 5.5 Map of a Sample Site Showing Systematic Sampling Locations 5-10

Figure 5.6 Method for Positioning Systematic Sample Locations in the Field 5-12

Figure 5.7 An Example Illustration of How to Choose an Exact Field Sampling Location from an Approximate Location 5-14

Figure 5.8 Subsampling and Sampling Across Depth 5-16

Figure 6.1 An Example of How to Group Sample Points from a Systematic Sample so that the Variance and Mean Can Be Calculated Using the Methodology for a Stratified Sample 6-24

TABLE OF CONTENTS

Page

Figure 6.2 Example of a Serpentine Pattern 6-25

Figure 8.1 Graphic Example of Sequential Testing.............................. 8-4

Figure 9.1 A Square Grid of Systematically Located Grid Points with Circular and Elliptical Hot Spots Superimposed 9-4

Figure 9.2 Grid Spacing and Ellipse Shape Definitions for the Hot Spot Search Table in Appendix A..9-5

Figure 10.1 An Example of an Empirical Variogram and a Spherical Variogram Model. 10-3

Figure 10.2 Contour Map of the Probability in Percent of Finding the Value of 1,000 ppm or a Larger Value 10-11

Figure 10.3 Contour Map of the Probability in Percent of a False Positive in the Remedial Action Areas and the 1,000 Contour Line 10-11

Figure 10.4 Contour Map of the Probability in Percent of a False Negative in the Remedial Action Areas and the 1,000 ppm Contour Line 10-11

Figure A.1 Power Curves for α = 1%.. A-12

Figure A.2 Power Curves for α = 5%.. A-13

Figure A.3 Power Curves for α = 10% A-14

Figure A.4 Power Curves for α = 25% A-15

Figure B.1 Example Worksheets: Parameters to Test in Each Sample Area and Map of the Site...B-2

Figure B.2 Example Worksheets: Sequence in Which the Worksheets are Completed B-3

TABLE OF CONTENTS

LIST OF TABLES

Page

Table 1.1 EPA guidance documents that present methodologies for collecting and evaluating soils data 1-5

Table 2.1 A diagrammatic explanation of false positive and false negative conclusions.. 2-5

Table 3.1 Points to consider when trying to choose among the mean, high percentile, or median 3-7

Table 3.2 Recommended parameters to test when comparing the cleanup standard to the average concentration of a chemical with chronic effects.. 3-9

Table 4.1 Where sample designs and analysis methods for soil sampling are discussed in this document............................ 4-7

Table 7.1 Selected information from Tables A.7-A.9 that can be used to determine the sample sizes required for zero or few exceedance rules associated with various levels of statistical performance and degrees of cleanup.................................7-14

Table 9.1 Selected references regarding the methodologies for identifying hot spots at waste sites.................................. 9-2

Table 9.2 Factors controlling the design of a hot spot search sampling plan 9-6

Table 10.1 Selected introductory and advanced references that introduce and discuss geostatistical concepts..................................10-5

Table 10.2 Introductory references for indicator, probability, and nonparametric global estimation kriging 10-10

Table 10.3 Selected geostatistical software.................................... 10-13

Table A.1 Table of t for selected alpha and degrees of freedom A-1

Table A.2 Table of z for selected alpha or beta A-2

Table A.3 Table of k for selected alpha, P_0, and sample size where alpha = 0.10 (i.e., 10%)...A-3

Table A.4 Table of k for selected alpha, P_0, and sample size where alpha = 0.05 (i.e., 5%)...A-4

Table A.5 Table of k for selected alpha, P_0, and sample size where alpha = 0.01 (i.e., 1%)...A-5

TABLE OF CONTENTS

Page

Table A.6 Sample sizes required for detecting a scaled difference tau of the mean from the cleanup standard for selected values of alpha and beta A-6

Table A.7 Sample size required for test for proportions with α = .01 and β = .20, for selected values of P_0 and P_1 .... A-7

Table A.8 Sample size required for test for proportions with α = .05 and β = .20, for selected values of P_0 and P_1 .... A-8

Table A.9 Sample size required for test for proportions with α = .10 and β = .20, for selected values of P_0 and P_1 .... A-9

Table A.10 Tables for determining critical values for the exact binomial test, with α = 0.01, 0.05, and 0.10 .... A-10

Table A.11 The false positive rates associated with hot spot searches as a function of grid spacing and hot spot shape A-11

TABLE OF CONTENTS

LIST OF BOXES

Page

Box 2.1 Estimating the Final Sample Size Required 2-15

Box 5.1 Steps for Generating Random Coordinates that Define Sampling Locations 5-4

Box 5.2 An Example of Generating Random Sampling Locations 5-5

Box 5.3 Calculating Spacing Between Adjacent Sampling Locations 5-7

Box 5.4 Locating Systematic Coordinates 5-9

Box 5.5 Alternative Method for Locating the Random Start Position for a Systematic Sample 5-11

Box 6.1 Calculating Sample Mean, Variance, Standard Deviation, and Coefficient of Variation 6-3

Box 6.2 An Alternate Approximation for $\hat{\sigma}$ 6-6

Box 6.3 Formulae for Calculating the Sample Size Needed to Estimate the Mean 6-7

Box 6.4 Example of Sample Size Calculations 6-8

Box 6.5 Example: Determining Sample Size for Testing the Mean Using the Power Curves 6-9

Box 6.6 Determining the Approximate Power Curve for a Specified Sample Size 6-10

Box 6.7 Computing the Upper One-sided Confidence Limit 6-11

Box 6.8 An Example Evaluation of Cleanup Standard Attainment 6-12

Box 6.9 Calculating the Proportion of the Volume of Soil 6-13

Box 6.10 Calculating Desired Sample Size for Each Stratum of a Stratified Random Sample 6-14

Box 6.11 An Example Sample Size Determination for a Stratified Sample 6-15

Box 6.12 Formula for the Mean Concentration from a Statified Sample 6-16

Box 6.13 Formula for the Standard Error from a Stratified Sample 6-17

Box 6.14 Formula for Degrees of Freedom from a Stratified Sample 6-17

LIST OF BOXES

Page

Box 2.1 Estimating the Final Sample Size Required 2-15

Box 5.1 Steps for Generating Random Coordinates that Define Sampling Locations 5-4

Box 5.2 An Example of Generating Random Sampling Locations 5-5

Box 5.3 Calculating Spacing Between Adjacent Sampling Locations 5-7

Box 5.4 Locating Systematic Coordinates 5-9

Box 5.5 Alternative Method for Locating the Random Start Position for a Systematic Sample 5-11

Box 6.1 Calculating Sample Mean, Variance, Standard Deviation, and Coefficient of Variation 6-3

Box 6.2 An Alternate Approximation for $\hat{\sigma}$ 6-6

Box 6.3 Formulae for Calculating the Sample Size Needed to Estimate the Mean 6-7

Box 6.4 Example of Sample Size Calculations 6-8

Box 6.5 Example: Determining Sample Size for Testing the Mean Using the Power Curves 6-9

Box 6.6 Determining the Approximate Power Curve for a Specified Sample Size 6-10

Box 6.7 Computing the Upper One-sided Confidence Limit 6-11

Box 6.8 An Example Evaluation of Cleanup Standard Attainment 6-12

Box 6.9 Calculating the Proportion of the Volume of Soil 6-13

Box 6.10 Calculating Desired Sample Size for Each Stratum of a Stratified Random Sample 6-14

Box 6.11 An Example Sample Size Determination for a Stratified Sample 6-15

Box 6.12 Formula for the Mean Concentration from a Statified Sample 6-16

Box 6.13 Formula for the Standard Error from a Stratified Sample 6-17

Box 6.14 Formula for Degrees of Freedom from a Stratified Sample 6-17

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA 230/02-89-042	2.	3. RECIPIENT'S ACCESSION NO. PB89 234959/AS
4. TITLE AND SUBTITLE METHODS FOR EVALUATING THE ATTAINMENT OF CLEANUP STANDARDS VOL I : SOILS AND SOLID MEDIA		5. REPORT DATE FEB 1989 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) BARNES JOHNSON, WESTAT (JOHN ROGERS, ADAM CHU, PAUL FLYER) DYNAMAC CORP (RICHARD DORSEY)		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS STATISTICAL POLICY BRANCH, US-EPA 401 M ST SW (PM-223) WASH. DC 20460		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO. * EPA Contract # 68-01-7359
12. SPONSORING AGENCY NAME AND ADDRESS Same as writing		13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This document provides regional project managers, onsite coordinators, and their contractors with sampling and analysis methods for evaluating whether a soils remediation effort has been successful. The verification of cleanup by evaluating a site relative to a cleanup standard or applicable and relevant or appropriate requirement (ARAR) is discussed in section 121 of the Superfund Amendments and Reauthorization Act (SARA). In section 121 of SARA the "attainment" of cleanup standards and ARARs is mentioned repeatedly. This manual, the first in a series, provides a technical interpretation of what sampling and data analysis methods are acceptable for verifying "attainment" of a cleanup standard in soils and solid media.

Statistical methods are emphasized because there is a practical need to make decisions regarding whether a site has met a cleanup standard in spite of uncertainty. The uncertainty arises because Superfund managers are faced with being able to sample and analyze only a small portion of the soil at the site yet having to make a decision regarding the entire site. Statistical methods are designed to permit this extrapolation from the results of a few samples to a statement regarding the entire site.

17. ** KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

STRIBUTION STATEMENT ELEASE UNLIMITED	19. SECURITY CLASS *(This Report)* UNC	21. NO. OF PAGES
	20. SECURITY CLASS *(This page)* UNC	22. PRICE

TABLE OF CONTENTS

Page

Box 6.15 Formula for the Upper One-sided Confidence Interval from a Stratified Sample 6-18

Box 6.16 An Example Illustrating the Determination of Whether the Mean from a Stratified Sample Attains a Cleanup Standard 6-19

Box 6.17 Estimating the Mean, the Standard Error of the Mean, and Degrees of Freedom When a Systematic Sample Is Treated as a Stratified Sample 6-24

Box 6.18 Formula for Upper One-sided Confidence Interval for the True Mean Contamination When a Systematic Sample Is Treated as a Stratified Sample 6-24

Box 6.19 Computational Formula for Estimating the Standard Error and Degrees of Freedom from Samples Analyzed in a Serpentine Pattern 6-26

Box 7.1 Illustration of Multiple Measurement Procedure for Reducing Laboratory Error 7-4

Box 7.2 Computing the Sample Size When Testing a Proportion or Percentile 7-6

Box 7.3 Example of How to Determine Sample Sizes When Evaluating Cleanup Standards Relative to a Proportion 7-7

Box 7.4 Calculating the Proportion Contaminated and the Standard Error of the Proportion 7-8

Box 7.5 Calculation of the Upper Confidence Limit on a Proportion Using a Large Sample Normal Approximation 7-9

Box 7.6 An Example of Inference Based on the Exact Test 7-11

Box 7.7 Computing the Sample Size for Stratum h 7-16

Box 7.8 Sample Size Calculations for Stratified Sampling 7-17

Box 7.9 Calculating an Overall Proportion of Exceedances and the Standard Error of the Proportion from a Stratified Sample 7-18

Box 7.10 Calculating the Upper Limit of the One-sided Confidence Interval on an Estimate of the Proportion 7-19

Box 7.11 Inference for Proportions Using Stratified Sampling 7-20

Box 7.12 Calculating the Sample Size Requirements for Tolerance Intervals 7-22

Box 7.13 Calculating Sample Size for Tolerance Intervals--Two Examples 7-23

TABLE OF CONTENTS

Page

Box 7.14 Calculating the Upper Tolerance Limit 7-24

Box 7.15 Tolerance Intervals: Testing for the 95th Percentile with Lognormal Data 7-25

Box 8.1 Defining Acceptance and Rejection Criteria for the Sequential Test of Proportions 8-6

Box 8.2 Deciding When the Site Attains the Cleanup Standard 8-7

Box 8.3 An Example of Sequential Testing 8-9

Box 8.4 Example of Sequential Test Using Grouped Samples 8-10

Box 9.1 Approximating the Sample Size When Area and Grid Interval Are Known 9-7

Box 10.1 Steps for Obtaining Geostatistical Software from EMSL-LV .. 10-12

AUTHORS AND CONTRIBUTORS

This manual represents the combined efforts of several organizations and many individuals. The names of the primary contributors, along with the role of each organization, are summarized below.

Dynamac Corporation, 11140 Rockville Pike, Rockville, MD 20852 (subcontractor to Westat) -- sampling, treatment, chemical analysis of samples. Key Dynamac Corporation staff included:

David Lipsky
Wayne Tusa
Richard Dorrler

EPA, OPPE, Statistical Policy Branch -- project management, technical input, peer review. Key EPA staff included:

Barnes Johnson

SRA Technologies, Inc., 4700 King Street, Suite 300, Alexandria, VA 22302, EPA Contract No. 68-01-7379, Task 11 -- editorial and graphics support, final report preparation. Key SRA staff included:

Marcia Gardner
Jocelyn Smith
Sylvia Burns
Lori Hidinger

Westat, Inc., 1650 Research Boulevard, Rockville, MD 20850, Contract No. 68-01-7359, Task 5 -- research, statistical procedures, draft report. Key Westat staff included:

John Rogers
Jill Braden
Paul Flyer
Ed Bryant
Adam Chu

EXECUTIVE SUMMARY

This document provides regional project managers, onsite coordinators, and their contractors with sampling and analysis methods for evaluating whether a soils remediation effort has been successful. The verification of cleanup by evaluating a site relative to a cleanup standard or applicable and relevant or appropriate requirement (ARAR) is discussed in section 121 of the Superfund Amendments and Reauthorization Act (SARA). In section 121 of SARA the "attainment" of cleanup standards and ARARs is mentioned repeatedly. This manual, the first in a series, provides a technical interpretation of what sampling and data analysis methods are acceptable for verifying "attainment" of a cleanup standard in soils and solid media.

Statistical methods are emphasized because there is a practical need to make decisions regarding whether a site has met a cleanup standard in spite of uncertainty. The uncertainty arises because Superfund managers are faced with being able to sample and analyze only a small portion of the soil at the site yet having to make a decision regarding the entire site. Statistical methods are designed to permit this extrapolation from the results of a few samples to a statement regarding the entire site.

The methods in this document approach cleanup standards as having three components that influence the overall stringency of the standard. The first component is the magnitude, level, or concentration that is deemed protective of public health and the environment. The second component of the standard is the sampling that is done to evaluate whether a site is above or below the standard. The final component is how the resulting data are compared with the standard to decide whether the remedial action was successful. All three of these components are important. Failure to address any of the three components can result in far less cleanup than desired. Managers must look beyond the cleanup level and explore the sampling and analysis that will allow evaluation of the site relative to the cleanup level.

For example, suppose that a cleanup level is chosen and that only a few samples are acquired. When the results are available, it is found that the mean of those samples is just below the cleanup level, and therefore, the site is judged as having been successfully remediated. Under this scenario, there may be a large chance that the average of the entire site, as opposed to the samples, is well above the cleanup level. Uncertainty was not considered, and therefore, there is a large chance that the wrong decision was made and the site-wide average is not below the cleanup level.

These concepts and solutions to the potential pitfalls are presented in a sequence that begins with an introduction to the statistical reasoning required to implement these methods. Then the planning activities are described; these require input from both nonstatisticians and statisticians. The statistical aspects of field sampling are presented. Finally, a series of methodological chapters are presented which consider the cleanup standard as: (1) an average condition; (2) a value to be rarely exceeded; (3) being defined by small discrete hot spots of contamination that should be found if present; or (4) broad areas that should be defined and characterized. A more detailed discussion of the document follows.

Chapter One introduces the need for the guidance and its application with risk-based standards, under various soils treatment alternatives, and in various parts of the Superfund program. Standards development and usage depends on certain factors, and the three categories of standards used by EPA are discussed: technology-based, background-based, and risk-based standards.

The statistical methods described in this manual are useful in various phases of treatment, testing, piloting, and full-scale implementation of various treatment technologies. In addition, the methods in this manual apply in various programmatic circumstances including both Superfund and Enforcement lead sites and removal actions.

Chapter Two addresses statistical concepts as they relate to the evaluation of cleanup attainment. Discussions of the form of the null and alternative hypothesis, types of errors, statistical power curves, and special data like less-than-detection-limit values and outliers are presented.

A site manager inevitably confronts the possibility of error in evaluating the attainment of the cleanup standard: is the site really contaminated because a few samples are above the standard? Conversely, is the site really "clean" because the sampling shows the majority of the samples to be within the cleanup standard? The statistical methods demonstrated in the guidance document allow decision making under uncertainty and valid extrapolation of information that can be defended and used with confidence to determine whether the site meets the cleanup standard.

The procedures in this guidance document favor protection of the environment and human health. If uncertainty is large or the sampling inadequate, these methods conclude that the sample area does not attain the cleanup standard. Therefore, the null hypothesis, in statistical terminology, is that the site does not attain the cleanup standard until sufficient data are acquired to prove otherwise.

Chapter Three discusses the steps in specifying attainment objectives. Definition of the attainment objectives is the first task in the evaluation of whether a site has attained a cleanup standard. Attainment objectives are not specified by statisticians, but must be provided by risk assessors, engineers, and soil scientists. Specifying attainment objectives includes specifying the chemicals of concern and the cleanup levels, as well as choosing the area to be remediated.

Chapter Four presents approaches to the design of remedial verification sampling and analysis plans. The specification of this plan requires consideration of how the environment and human health are to be protected and how the sampling and analysis are to achieve adequate precision at a reasonable cost.

Sampling designs considered in this guidance document are random sampling, stratified sampling, systematic sampling, and sequential sampling. Differences in these approaches, including advantages and disadvantages, are both discussed and graphically displayed. With any plan, the methods of analysis must be consistent with the sample design.

A primary objective of the analysis plan involves making a decision regarding how to treat the applicable cleanup standard. For example, is the cleanup standard a value that should rarely be exceeded (1 or 5 percent of the time) at a remediated site? Or, alternatively, should there be high confidence that the mean of the site is below the cleanup standard? Should there be no hot spots with concentrations in excess of the cleanup standard? Or should the analysis plan employ a combination of these criteria.

Chapter Five discusses the statistical aspects of field sampling procedures. The procedures used to establish random and systematic sample locations are discussed. In addition to selecting sampling locations, the advantages and disadvantages of methods for subsampling across depth are discussed and illustrated. Three approaches presented are depth discrete sampling, compositing across depth, and random sampling across depth.

Chapter Six describes procedures for determining whether there is confidence, based on the results of a set of samples, that the mean concentration of the contaminant in a sample area is less than the cleanup standard. Basic formulas are given and used in examples to illustrate the procedures. The primary point is that to ensure with confidence that the site mean is below the cleanup standard, the sample mean must be well below the cleanup standard by a distance determined by a confidence limit.

The following topics--determination of sample size; calculation of the mean, standard deviation, and confidence interval; and deciding if the sample area attains the cleanup standard--are discussed for these three sampling plans:

- Simple random sampling;
- Stratified random sampling; and
- Systematic sampling.

Chapter Seven presents several approaches that allow evaluation of whether a specified proportion or percentage of soil at a hazardous waste site is below the cleanup standard. The methods described apply if there is interest in verifying that only a small proportion or percentage of the soil at the site exceeds the cleanup standard.

One way to implement these methods is to use simple exceedance rules. A sample size and number of exceedances are specified that coincide with an acceptable level of certainty and level of cleanup. If the prespecified number of samples is obtained and the number of exceedances is less than or equal to the allowed number of exceedances the site is judged clean. If there are more exceedances than allowed then cleanup cannot be verified. The more exceedances allowed, the more soil samples that need to be collected to maintain the statistical performance of the method.

Chapter Eight deals with sequential sampling as a method for testing percentiles. Unlike the fixed-sample-size methods discussed in the two previous chapters, with sequential sampling, a statistical test is performed after each sample or small batch of samples is collected and analyzed. The test then makes one of three decisions: the site has attained the cleanup standard, the site has not attained the cleanup standard, or select another sample. Sequential sample findings can respond quickly to very clear or very contaminated sites and therefore offer cost savings. Although these procedures yield a lower sample size on the average than that for fixed-sample-size procedures, in order to be practical, they require "rapid turn-around" laboratory methods.

Chapter Nine illustrates the design of sampling plans to search for hot spots. The conclusions that can be drawn regarding the presence or absence of hot spots are discussed. Hot spots are generally defined as relatively small, localized, elliptical areas with contaminant concentrations in excess of the cleanup standard. Tables are provided to help determine grid spacing and detect hot spots of various sizes with different probabilities.

Chapter Ten discusses the use of geostatistical methods, which provide a method for mapping spatial data that enables both interpolation between existing data points and a method for estimating the precision of the interpolation. Geostatistical applications are described as a two-step process. First, the spatial relationship is modeled as a variogram and then the variogram is used by a kriging algorithm to estimate concentrations at points that were not sampled. Indicator and probability kriging are most useful for remedial verification purposes.

Geostatistical methods have many applications in soil remediation technology, especially when the extent of contamination needs to be characterized. This chapter includes guidance to help decide whether geostatistical data analysis and evaluation methods are appropriate for use with soils remediation activities that involve removal, homogenization, and flushing.

Before being applied the kriging techniques will require further study on the part of the user. Reference documents are listed. Because kriging cannot be conveniently or practically implemented without a computer and the appropriate software, a first-level familiarity with the methodology along with use of a software package is desirable to explore example applications and data sets. EPA has developed the first version of a geostatistical software package which can be obtained by following instructions at the end of Chapter 10.

1. INTRODUCTION

Congress revised the Superfund legislation in the Superfund Amendments and Reauthorization Act of 1986 (SARA). Among other provisions of SARA, section 121, Cleanup Standards, discusses criteria for selecting applicable and relevant or appropriate requirements (ARARs) and includes specific language that requires EPA cleanups to **attain** ARARs.

Neither SARA nor EPA regulations or guidances specify how to determine attainment or verify that the cleanup standards have been met. This document offers procedures that can be used to determine whether, after a remediation action, a site has attained an appropriate cleanup standard.

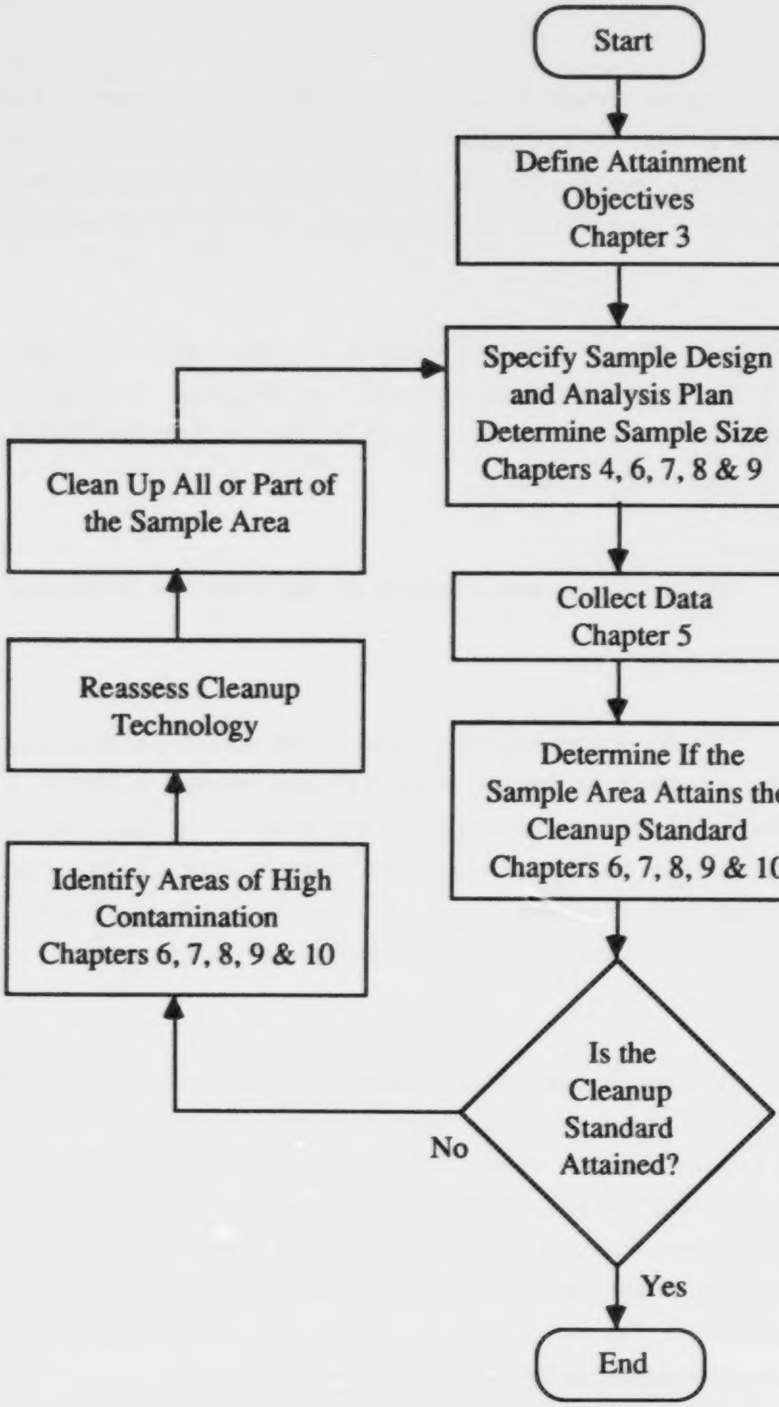
1.1 General Scope and Features of the Guidance Document

1.1.1 Purpose

This document describes methods for testing whether soil chemical concentrations at a site are statistically below a cleanup standard or ARAR. If it can be reasonably concluded that the remaining soil or treated soil at a site has concentrations that are statistically less than relevant cleanup standards then the site can be judged protective of human health and the environment. Figure 1.1 shows the steps involved in this evaluation which requires specification of attainment objectives, sampling protocols, and analysis methods.

For example, consider the situation where several samples were taken. The results indicate that one or two of the samples exceed the standard: How should this information be used to decide whether the standard has been attained? Some possible considerations include: the mean of those samples could be compared with the standard; the magnitude of the two sample values that are larger than the standard might be useful in making a decision; or the area where the two large sample values were obtained might provide some insight. The following factors are important in reaching the decision as to whether a cleanup standard has been attained:

Figure 1.1 Steps in Evaluating Whether a Site Has Attained the Cleanup Standard


Start
Define Attainment Objectives Chapter 3
Specify Sample Design and Analysis Plan Determine Sample Size Chapters 4, 6, 7, 8 & 9
Clean Up All or Part of the Sample Area
Collect Data Chapter 5
Reassess Cleanup Technology
Determine If the Sample Area Attains the Cleanup Standard Chapters 6, 7, 8, 9 & 10
Identify Areas of High Contamination Chapters 6, 7, 8, 9 & 10
Is the Cleanup Standard Attained?
No
Yes
End

- The spatial extent of the sampling and the size of the sample area;
- The number of samples taken;
- The strategy of taking samples; and
- The way the data are analyzed.

Simply to require that a Superfund site be cleaned until the soil concentration of a chemical is below 50 mg/kg is incomplete. Statements suggesting that the site will be remediated until the soil concentration of a chemical is 50 mg/kg reveal little in terms of the environmental results anticipated, the future exposure expected, the resultant risk to the local population, or the likelihood that substantial contamination will remain after a decision is made that the site has been fully remediated. A specific sampling and data analysis protocol must accompany the risk-based standard for the standard to be meaningful in terms of benefit or actual risk.

This document does not attempt to suggest which standards apply or when they apply (i.e., the "How clean is clean?" issue). Other Superfund guidance documents (e.g., USEPA,1986c and USEPA, 1986d) perform that function.

1.1.2 Intended Audience and Use

Management/supervisory personnel will find the executive summary and introductory chapters useful. However, this manual is intended primarily for Agency personnel (primarily onsite coordinators and regional project managers), responsible parties, and their contractors who are involved with monitoring the progress of soils remediation at Superfund sites. Although selected introductory statistical concepts are reviewed, the document is directed toward readers that have had prior training or experience applying quantitative methods.

This document discusses data analysis and statistical methods for evaluating the effectiveness of Superfund remedial actions. However, there are many other technical aspects to this problem. Input from soil scientists, engineers, geologists, hydrologists, geochemists, and analytical chemists is essential. There must be dialogue among this group, including the statistician, so that each member understands and considers the point of view of the others. It is only through collaboration that an effective evaluation scheme can be developed to measure the effectiveness of a remedial action.

This document does not intend to address the issues that the other members of the team specialize in such as:

- Soil sample acquisition protocols;
- Areas of the vadose zone of concern under different situations;
- The influence of soil chemistry;
- Waste types based on industrial processes;
- Leaching procedures that approximate the expected weathering processes and risk assessment assumptions;
- Chemical analysis methods useful given particular soils matrices; or
- Approaches to soils remediation.

Table 1.1 lists other relevant EPA guidance documents on sampling and evaluating soils and solid media that apply to both the statistical and other technical components of a sampling and analysis program.

The selection of statistical methods for use in assessing the attainment of cleanup standards depends on the characteristics of the data. In soils, concentrations of contaminants change relatively slowly, with little variation from season to season. In ground water, the number of measurements available for spatial characterization is limited and seasonal patterns may exist in the data. As a result of these differences, separate procedures are recommended for the differing problems associated with soils and solid media, and ground water, surface water, and air. These media will be addressed in separate volumes.

Table 1.1 EPA guidance documents that present methodologies for collecting and evaluating soils data

Title	Sponsoring Office	Date	ID Number
Preparation of Soil Sampling Protocol: Techniques and Strategies	EMSL-LV ORD	August 1983	EPA 600/ 4-83-020
Soil Sampling Quality Assurance User's Guide	EMSL-LV ORD	May 1984	EPA 600/ 4-84-043
Verification of PCB Spill Cleanup by Sampling and Analysis	OTS OPTS	August 1985	EPA 560/ 5-85-026
Guidance Document for Cleanup of Surface Impoundment Sites	OERR OSWER	June 1986	OSWER DIRECTIVE 9380.0-6
Test Methods for Evaluating Solid Waste	OSW OSWER	November 1987	SW-846
Draft Surface Impoundment Clean Closure Guidance Manual	OSW OSWER	March 1987	OSWER DIRECTIVE 9476.0-8.C
Data Quality Objectives for Remedial Response Activities: Development Process	OERR OSWER	March 1987	EPA 540/ G-87/003
Data Quality Objectives for Remedial Response Activities: Example Scenario RI/FS Activities at a Site with Contaminated Soils and Ground Water	OERR OSWER	March 1987	EPA 540/ G-87/004

It must be emphasized that this document is intended to provide flexible guidance and general direction. This manual is not a regulation and should not be imposed as a regulation. Finally, this document should not be used as a "cookbook" or a replacement for engineering judgment.

1.1.3 Bibliography, Glossary, Boxes, Worksheets, Examples, and References to "Consult a Statistician"

The document includes a bibliography which provides a point of departure for the user interested in further reading. There are references to primary textbooks, pertinent journal articles, and related guidances.

The glossary is included to provide short, practical definitions of terminology used in the manual. The glossary does not use theoretical explanations or formulae and should not be considered a replacement for more complete discussions in the text or alternative sources of information.

Boxes are used throughout the document to separate and highlight calculation methods and example applications of the methods. A listing of all boxes and their page numbers is provided on pages xii - xiv.

A series of worksheets is included to help organize calculations. Reference to the pertinent sections of the document appears at the top of each worksheet.

Example data and calculations are presented in the boxes and worksheets. The data and sites are hypothetical, but elements of the examples correspond closely to actual sites.

Finally, the document often directs the reader to "consult a statistician" when more difficult and complicated situations are encountered. A directory of Agency statisticians is available from the Statistical Policy Branch (PM-223) at EPA Headquarters.

1.2 A Categorization Scheme for Cleanup Standards

Superfund remediations require standards for assessing the success and completion of the cleanup. The criteria for choosing the type of standard and setting the magnitude of the standard come from different sources, depending on many factors including the nature of the contamination, negotiations with potentially responsible parties, and public comment on alternatives identified by EPA.

Many programs throughout EPA use numerical standards variously described as ARARs, concentration limits, limitations, regulatory thresholds, action levels, and criteria. These standards are often expressed as concentration measures of chemicals or chemical indicators. Standards development and usage depends on the media to which the standard applies, the data used to develop the standard, and the manner of evaluating compliance with the standard. The following discussion categorizes the standards used by EPA and compares the features of each category.

1.2.1 Technology-Based Standards

Technology-based standards are developed for the purpose of defining the effectiveness of pollution abatement technology from an engineering perspective. For example, waste water treatment plants operating under the National Pollution Discharge Elimination System (NPDES) must be designed and operated under a numerically prescribed level of technological performance depending on the particular industrial category. Technology-based standards such as the NPDES standards are developed and applied using statistical methods that consider variability in the operation of the treatment system. The likelihood of exceeding the standards is rare if the technology is installed and operated properly. Often Superfund sites require the installation of waste water treatment systems and compliance with NPDES standards.

1.2.2 Background-Based Standards

Background-based standards are developed using site-specific background data. An example is the background ground water concentration standards that hazardous waste land disposal facilities use under Resource Conservation and Recovery Act (RCRA) permits. The background data are used to establish a standard for the facility, which accounts for the presence of any existing contamination hydraulically upgradient of the facility. Background standards are applied on a site-specific basis, but because they are developed using statistical methodologies, the standards can be associated with a known false positive and false negative rate.

1.2.3 Risk-Based Standards

A third class of standards, risk-based standards are developed using risk assessment methodologies. Chemical-specific ARARs adopted from other programs often include at least a generalized component of risk. However, risk standards may be specific to a site, developed using a local endangerment evaluation.

Risk-based standards are expressed as a concentration value. However, cleanup standards based on risk as applied in the Superfund program are not associated with a standard method of interpretation when applied in the field. Therefore, risk-based standards, when applied in the field, do not consider false positive and false negative errors. Although statistical methods are used to develop elements of risk-based standards, the estimated uncertainties are not carried through the analysis or used to qualify the standards for use in a field sampling program. Even though risk standards are not accompanied by measures of uncertainty, field data, collected for the purpose of representing the entire site and validating cleanup, will be uncertain. This document allows decision making regarding site cleanup by providing methods that statistically compare risk standards with field data in a scientifically defensible manner that allows for uncertainty.

1.3 Use of this Guidance in Superfund Program Activities

Standards that apply to Superfund activities normally fall into the third category of risk-based standards. There are many Superfund activities where risk-based standards might apply. The following discussion provides suggestions for using the methods described in this document in the implementation and evaluation of Superfund activities.

1.3.1 Emergency/Removal Action

Similar to the guidance regarding sampling strategies associated with PCB spills (USEPA, 1985), cleanup activities associated with the methods in this document will be useful for circumstances that are encountered during emergency cleanups and removals. In many cases, because of the time, safety, and exposure constraints associated with emergency activity, initial cleanup will focus on areas visually or otherwise known to be contaminated. The methods described in this document will, however, be most useful in verifying the initial cleanup of obvious contamination.

1.3.2 Remedial Response Activities

The objective of remediation is to ensure that release of and exposure to contaminants is curtailed. Remedial efforts are normally long-term and require diverse, innovative technology. As discussed in section 1.4, soil or solid media remediation can be addressed using a variety of technologies. Numerical standards are used to define the degree of curtailment. The methods described in this document can help to evaluate the utility of the remediation technology in treating contaminants with respect to a particular numerical standard.

1.3.3 Superfund Enforcement

The methods described in this document will also be useful for providing more technically exacting negotiations, consent decree stipulations, and responsible party oversight. Questions such as "How much is enough?" and "When can I stop cleaning?" are constantly emerging at the enforcement negotiation table. More specific questions such as "How much should you sample?", "What sampling pattern or method of sampling design should be applied?" and "How can I minimize the chance of saying the site is still dirty when it is basically clean?" are addressed here, as well as the ultimate question: "How do I know when the standard has been attained at the entire site, knowing that the decision is based on a body of data that is incomplete and uncertain?"

1.4 Treatability Studies and Soils Treatment Technologies

In addition to discussing the methods described in this document and their relationship to aspects of the Agency's Superfund program, it is also important to discuss how the methods will function when applied in treatability studies and under various soils treatment technologies.

1.4.1 Laboratory/Bench-Scale Treatability Studies

Feasibility studies often include small bench-scale laboratory evaluations of how various treatment agents and concentrations of agents will perform. Suppose that the contaminant and soil characteristics at the site indicate that two fixation media offer a promising remediation approach. A treatability study examining several concentrations of the two media is proposed.

Under this scenario, the methods described in this manual could be applied to the sampling program used to obtain soils material for the treatability study. Treatability studies require "worst case" material--that is, soils with the highest concentrations or with the most tightly bound contaminants. Therefore, "worst case" sample areas within the site must be delineated, using data from prior remedial investigations. Once the "worst case" sample area is defined, the soils can be sampled as described in this manual, the treatability study executed, and the resulting data analyzed using the methods described in this document to examine whether the method has sufficiently treated the soil to allow attainment of the relevant cleanup standard.

1.4.2 Field/Pilot-Scale Treatability Studies

Once the feasibility study establishes an effective approach to treatment, it may be implemented as an onsite pilot using the chemical/physical/biological remedy with construction-scale onsite machinery. The approach favored in the bench-scale laboratory experiment may be chosen if the cost is reasonable. Machinery such as cement mixers, soil washers, soil mixing augers, incinerators, vacuum extraction manifold networks, or infiltration or injection systems are used in a pretest fashion. With an associated monitoring program, the methods in this guidance can be applied to determine whether the method will attain the desired level of cleanup.

The primary difference between the laboratory testing results and those obtained from field-scale pilot application is that far greater variability will be encountered in the onsite pilot. Unless the treatment method is exceptionally effective relative to risk-based standards in the laboratory, the variability encountered in the field may obscure the treatment's effectiveness. This document guides the user to methods that will help in such a situation. In addition, if a reasonable sampling program is conducted at the pilot-scale, these data can be used to estimate sample sizes for the sampling program associated with the full-scale implementation of the technology.

1.4.3 Soils Treatment by Chemical Modification

Soils are often treated by chemical fixation or stabilization. This technology uses a cement or grout-like material mixed with the contaminated soil or sediment. Once the mixture reacts, it solidifies, and contaminants are retained in the matrix and resist leaching. When this technology is used, the methods in this manual can be applied, keeping in mind, however, the following caveats.

Once the material has solidified onsite, it cannot be sampled easily. The ability to stabilize the site may be compromised if cores were obtained throughout the area. In addition, the resulting monolith may be capped, which would restrict access to the solidified matrix. Because it cannot be sampled after fixation, monitoring plans should be developed before the mixing occurs. The sampling could occur by taking samples at randomly located positions across the site and then pouring cylindrical casts of the material immediately after it is mixed prior to setup. Enough casts must be obtained for the initial evaluations of the site and for monitoring the aging process of the stabilized material. During analysis, concentrations are measured in leachate obtained from an accepted extraction procedure. Evaluation of the leachate from the casts allows determination of whether the lithified material attains or continues to attain the relevant cleanup standard.

1.4.4 Soils Treatment by In Situ Removal of Contaminants

Several soil treatment technologies, including vacuum extraction, soils leaching, and bioremediation, remove the contaminants without massive soil movement. The efficacy of these systems can be evaluated using the methods in this document, with the exceptions noted below.

Vacuum extraction is used to remove volatile compounds. Ambient air is drawn down through the soil into a well network and then into an adjustable manifold system attached to a vacuum pump. Air is then sent through carbon columns to remove the volatile compounds.

Soils leaching technologies are generally designed to extract contaminants that are water soluble. Soils leaching also relies on a network of wells attached to a manifold system. The system includes infiltration areas where aqueous solutions are allowed to recharge into the soils system. A pumping system is attached to the manifold and the water, after migration from the infiltration area to each well, is extracted and sent to a waste water treatment system.

Soils bioremediation can be used to degrade contaminants. Microorganisms use the contaminant as an energy source. One or more injection wells introduce water possibly enriched with oxygen, nutrients, microorganisms, or other essential growth-promoting materials. The injection wells are installed on one side of the contaminated area and monitoring wells are installed in various patterns throughout and possibly beyond the area of contamination. Again, a manifold system might be used for injection or sampling, and extraction wells may be used to direct or improve water movement.

With these technologies, something other than direct soils sampling may be used to evaluate effectiveness of the remediation, for example mass balance differencing. In this case, the methods herein may not always apply. However, monitoring of the soil relative to a risk-based cleanup standard is the most direct and protective measure of any soils cleanup technology.

Another concern is that when these systems are in place, the above-ground or slightly buried piping network will restrict the access of soils sampling equipment. For example, vehicle-deployed augers may not be able to reach certain areas. Engineering specifications should call for easy disassembly whenever possible. In cases where this is not possible, the guidance can still be applied after exclusion of certain soil areas because of inaccessibility.

A third consideration is that, after implementing the soil remediation technology, the soils concentration profile may begin to take on a regular spatial pattern. This occurs because removal wells are often arranged in a grid pattern and each well has a zone of influence where the concentrations have been reduced substantially. The result is a series of areas with high and low concentrations across the site. As discussed in the sampling chapter, under these circumstances systematic sampling should not be used because all or many of the samples may be located in areas with high or low concentrations. Random sampling is recommended to avoid this problem.

A final concern is that the soils system must be at steady state during the sampling program. This requires shutting down the extraction process and allowing the system to return to its original balance. This may take some time depending on characteristics of the system. In some cases when progress is being measured over time, methods pertaining to ground water in Volume 2 of this series might be more appropriate.

1.4.5 Soils Treatment by Incineration

Soils incineration involves the burning of soils in a furnace at high temperatures to degrade the contaminants into a nontoxic form. The product of the incineration is an ash. If questions arise as to whether the ash material contains chemicals in excess of applicable standards, then this manual might be useful. Sampling will have to be designed based on site-specific circumstances. If the treatment is highly effective and uniform, only a few samples may be necessary to verify effectiveness. However, if the standard is quite low and the measurement technology is variable at low concentrations, more samples may be required.

1.4.6 Soils Removal

In the soils removal approach to site cleanup, soils are permanently or temporarily removed from the site. Sampling must be done to verify that enough soil has been removed, and to ensure that clean soil is not needlessly removed. Under the circumstances associated with soils removal, there is no homogenization of the soil through a fixation process or artificial regularity to the soil profile caused by local extraction. In this case, geostatistical applications (Chapter 10) are useful for characterizing the contaminant profile. A new concentration profile can be estimated with each succeeding layer that is removed. In addition, geostatistical applications can help to identify hot spots that should be removed and sampling and analysis to detect hot spots might be useful (Chapter 9).

Finally, the simpler, more conventional evaluation methods that comprise the bulk of this manual can also be used. Exner et al. (1985) describe an application of these evaluation methods to a soils removal scenario at a Superfund site with dioxin contamination.

1.4.7 Soils Capping

A final category of soils remediation is to cap a site with impermeable layers of clay and synthetic membranes. This prevents surface water from recharging to the ground water through contaminated soils. Often caps are added as an additional measure in conjunction with other approaches. The methods in this document can be used to determine whether caps have met an engineering specification. For example, if the cap is intended to be constructed with no more than a 10^{-7} cms/sec permeability, samples might be obtained to document that permeability has been attained. Sampling may be difficult because it might disturb the integrity of the cap; however, it is possible that a pilot-scale procedure could be implemented to verify attainment of the standard.

1.5 Summary

This document deals with statistical methodology and procedures for use in assessing whether, after remediation, the treated soil or remaining soil attain the cleanup standards that are protective of public health and the environment as required by section 121 of SARA.

Use of the document is intended primarily for Agency personnel, responsible parties, and contractors who are involved with monitoring the progress of soils and remediation at Superfund sites. Although selected introductory statistical concepts are reviewed, the document is directed toward users having prior training or experience in applying quantitative methods.

Important factors in determining whether a cleanup standard has been attained are:

- The spatial extent of the sampling and the size of the sample area;
- The number of samples taken;

- The strategy of taking samples; and
- The way the data are analyzed.

The three types of EPA cleanup standards are technology-based standards, background-based standards, and risk-based standards. Superfund activities usually employ risk-based standards. By providing methods that statistically compare risk standards with field data in a scientifically defensible manner that allows for uncertainty, this document allows decision making regarding site cleanup. The statistical methods can be applied to the implementation and evaluation of:

- Emergency/removal action,
- Remedial response activities, and
- Superfund enforcement.

Also discussed are the functions of the statistical methods described in the document in the context of a variety of treatability studies and soils treatment technologies.

2. INTRODUCTION TO STATISTICAL CONCEPTS AND DECISIONS

When it comes to verifying cleanup, suppose that no exceedances of the cleanup standard are to be allowed. In that case, one of the most frequently asked questions regarding the use of statistical techniques in the evaluation of cleanup standards is:

Why should I use statistical methods and complicate the remedial verification process?

Allowing no exceedances of a standard is a perfectly acceptable decision rule to use. In fact, that simple rule is a statistical procedure because errors are possible. However, there is a chance that no exceedances will be discovered, yet a substantial portion of the site is above the cleanup standard. This is clearly not a desirable environmental result. With small sample sizes the chance of missing contamination is greater than with larger sample sizes. This is intuitive; the more you search for contamination and do not find it, the more confident you become in your conclusion that the site is clean.

Alternatively, consider the situation where a reasonable number of samples is taken and one sample exceeds the cleanup standard. In this case, you would conclude that the site continues to be dirty under the no exceedance rule. However, the problem is that this conclusion may be in error. Either laboratory error occurred or some rare and insignificant parcel of contamination could have been discovered. Revisiting the remedial method after many years or dollars of implementation is not reasonable because of the possibility that an error was made. As sample sizes are increased, the chances of finding one of the few obscure samples above the cleanup standard increases. How can you balance the two sets of possibilities: the chance that the site is contaminated even when the sampling shows attainment of the cleanup standard, and the chance of contamination when the majority of samples taken show the site to be clean?

The answer is to evaluate the potential magnitude of these two errors and balance them using the statistical strategies described in this manual. Statistical methods perform a powerful and useful function--they allow extrapolation from a set of samples to the entire site in a scientifically valid fashion.

Consider the following circumstance. The surface layer of soil from the bottom of a 4-hectare lagoon at a Superfund site will be sampled using cores with a 4-cm area. Given the size of the core and lagoon there will be approximately 10 million sample locations; however, concentration measurements will only be made on 100 of the 10 million. Statistical sampling and analysis methods provide an approach for choosing which 100 of the 10 million locations to sample so that valid results can be presented and statements can be made regarding the characteristics of the 10 million potential samples or the entire site.

Clearly, because of the extrapolation exercise, the statements or inferences regarding the 10 million sample locations have uncertainty. Statistical methods enable estimation of the uncertainty. Without the statistical methods, uncertainty still exists; but the uncertainty cannot be estimated validly.

This chapter will elaborate on statistical concepts and their specific application to the evaluation of cleanup standards. Statistical concepts such as the form of the null and alternative hypothesis, types of errors, statistical power, and handling peculiar data structures like less-than-detection-limit values and outliers are discussed to promote understanding. However, it is not necessary to read this chapter to apply the methods in this manual.

2.1 Hypothesis Formulation and Uncertainty

With any statistical procedure, conclusions will vary depending on which soil sample locations are selected. Therefore, based on the data collected, the investigator may conclude that:

- The site attains the cleanup standard;
- The site does not attain the cleanup standard; or
- More information is required to make a decision with a specified level of confidence.

When the results of the investigation are uncertain, the procedures in this guidance document favor protection of the environment and human health and conclude that the sample area does **not** attain the cleanup standard. In the statistical terminology applied in this document, the null hypothesis is that the site does not attain the cleanup standard. The null hypothesis is assumed to be true unless substantial evidence shows that it is false. Let ϕ represent the true (but unknown) value of a particular soil property, such as the mean concentration of a specified chemical over the entire site. The null hypothesis is:

$$H_0: \phi \geq \text{Cleanup Standard (CONTAMINATED or DIRTY)},$$

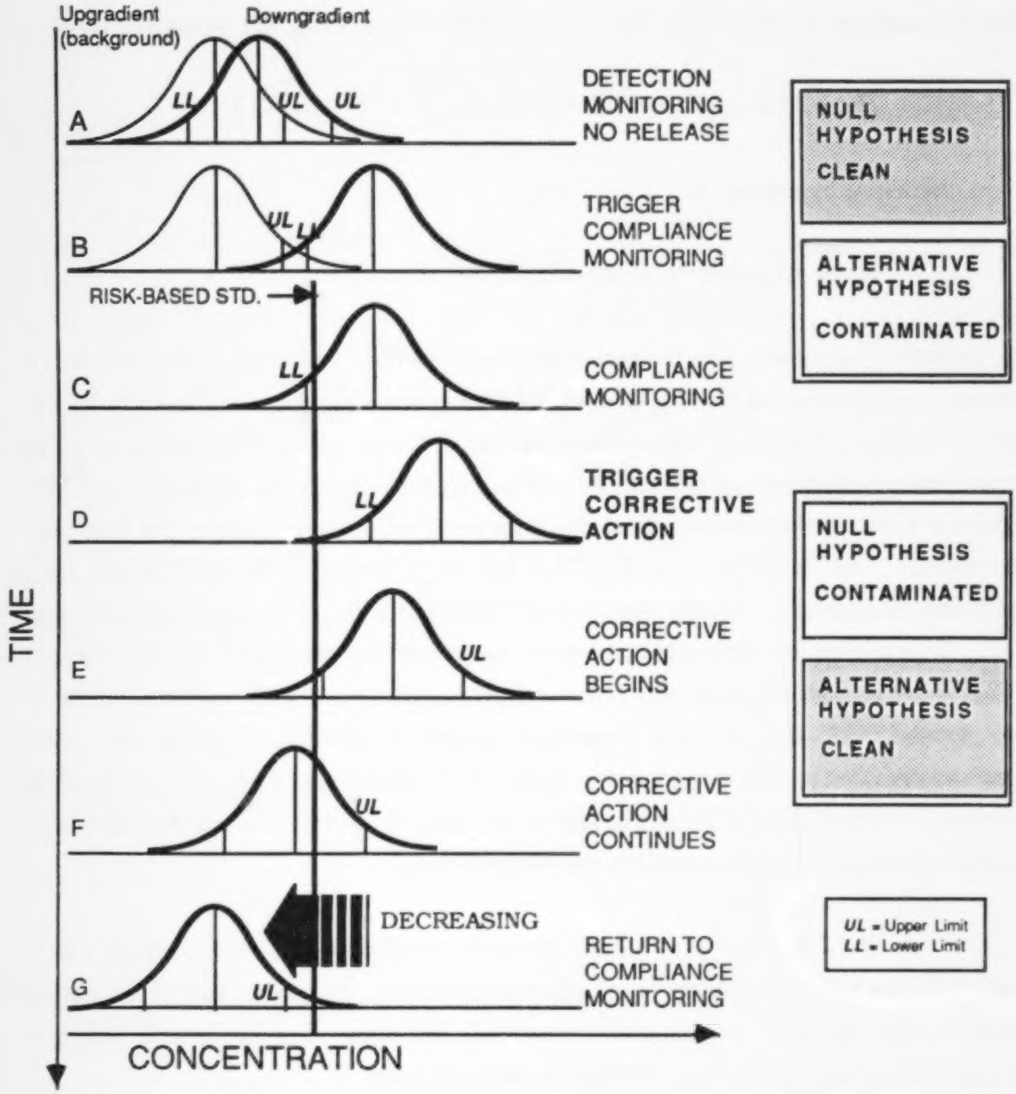
and the alternative hypothesis is:

$$H_1: \phi < \text{Cleanup Standard (CLEAN)}.$$

This document describes how to gather and analyze data that will provide evidence necessary to contradict the null hypothesis and demonstrate that the site indeed attains the cleanup standard. Figure 2.1 shows how the null and alternative hypothesis change as contamination is detected and subsequently corrected. This illustration specifically pertains to ground water evaluations for land disposal facilities operating under the Resource Conservation and Recovery Act (RCRA), but the concept is similar for the soils contamination situation. Initially, the the null hypothesis is that there is no contamination (A-C). Once a statistical demonstration can be made that the downgradient concentrations are first above background-level concentrations (B) and also above a relevant action limit or other standard (D), then the null hypothesis is that the site is contaminated. Most Superfund sites that require cleanup are in the situation described by D-E. The site must, at that point, be remediated (E,F) and proven to be clean (G) before the null hypothesis as described above can be rejected and the site declared clean.

If the null and alternative hypothesis described above were reversed, then a situation similar to C would designate a satisfactory cleanup. As can be seen by comparing C with G, the improper specification of the null and alternative hypothesis during a corrective action can result in very different levels of cleanup.

Figure 2.1 A Statistical Perspective of the Sequence of Ground Water Monitoring Requirements Under RCRA


Upgradient (background)
Downgradient
LL
UL
UL
A
DETECTION MONITORING NO RELEASE
NULL HYPOTHESIS CLEAN
B
UL
LL
TRIGGER COMPLIANCE MONITORING
ALTERNATIVE HYPOTHESIS CONTAMINATED
RISK-BASED STD.
C
LL
COMPLIANCE MONITORING
D
LL
TRIGGER CORRECTIVE ACTION
NULL HYPOTHESIS CONTAMINATED
TIME
E
UL
CORRECTIVE ACTION BEGINS
ALTERNATIVE HYPOTHESIS CLEAN
F
UL
CORRECTIVE ACTION CONTINUES
DECREASING
G
UL
RETURN TO COMPLIANCE MONITORING
UL = Upper Limit
LL = Lower Limit
CONCENTRATION

(Notice that until contamination above a risk standard is documented (D) the null hypothesis is that the facility is clean. Once the facility has been proven to be in exceedance of a health criteria then the null hypothesis is that the facility is contaminated until proven otherwise (G).)

When specifying simplified Superfund site cleanup objectives in consent decrees, records of decision, or work plans, it is extremely important to say that the site shall be cleaned up until the sampling program indicates with reasonable confidence that the concentrations of the contaminants at the entire site are statistically less than the cleanup standard. This prescription will result in the site being designated clean only after a situation similar to G is observed. However, attainment is often wrongly described by saying that concentrations at the site shall not exceed the cleanup standard. This second prescription can result in a situation similar to C being designated as clean.

As discussed in the introduction to this chapter, variation in sampling and lab analysis introduces uncertainty into the decision concerning the attainment of a cleanup standard. As a result of the uncertainty and the null/alternative hypothesis arrangement discussed above, the site can be determined clean when, in fact, it is not, resulting in a **false positive** decision (or Type I error). The converse of a false positive decision is a **false negative** decision (or Type II error), the mistake of saying the site needs additional cleanup when, in fact, it meets the standard. The Greek letter **alpha** (α) is used to represent the probability of a false positive decision and **beta** (β) is used to represent the probability of false negative decision. The definitions above are summarized in Table 2.1.

Table 2.1 A diagrammatic explanation of false positive and false negative conclusions

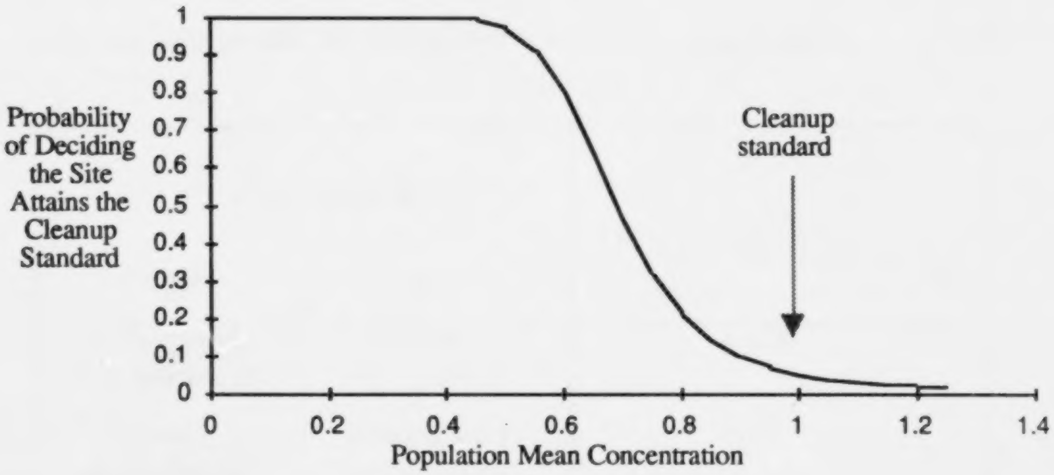
Decision based on the sample data is:	The true condition is:	
	Clean	Dirty
Clean	Correct Power (1 - β)	False positive (Probability is α)
Dirty	False negative (Probability is β)	Correct Certainty (1 - α)

It can be seen that if both α and β can be reduced, the percent of time that the correct decision will be made will be increased. Unfortunately, simultaneous reduction usually can be achieved only by increasing sample size, which may be expensive.

2.2 Power Curves as a Method of Expressing Uncertainty and Developing Sample Size Requirements

The probability of declaring the sample area clean will depend on the population mean concentration. If the true population mean is above the cleanup standard the sample area will rarely be declared clean (this will only happen if the mean of the particular set of samples is by chance well below the cleanup standard). If the population mean is much smaller than the cleanup standard, the sample area will almost always be judged clean. This relationship can be demonstrated by Figure 2.2. The figure illustrates a power curve that shows the probability of deciding that the site attains the cleanup standard on the vertical axis and the true, but always unknown, population mean concentration on the horizontal axis.

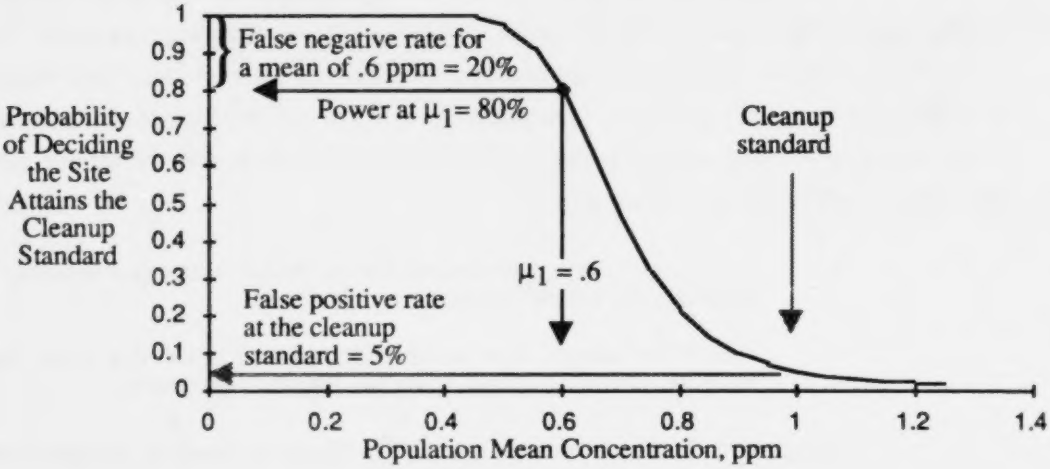
Figure 2.2 Hypothetical Power Curve



If the population mean concentration in the sample area is equal to or just above the cleanup standard (i.e., does not attain the cleanup standard), there is still a small 5-percent probability of declaring the sample area clean; this is the false positive rate denoted by α.

If the population mean is equal to 0.6 ppm (i.e., attains the cleanup standard of 1.0 ppm), the probability of declaring the sample area clean is 80 percent. Conversely the probability of declaring the site dirty, given that it is actually clean, is 20 percent. This is the false negative rate for a population mean of 0.6 ppm. Note that the probability of declaring the site clean changes depending on the population mean. These false positive and false negative rates are shown in Figure 2.3.

Figure 2.3 Hypothetical Power Curve Showing False Positive and False Negative Rates



The following items specify the shape and location of the power curve:

- The population coefficient of variation;
- The method of sample selection (the sampling plan);
- The statistical test to be used;

- The false positive rate; and
- The sample size.

In summary, there are two important uses of power curves. The first is to further facilitate understanding of the concept that, although the site may actually be clean, a set of samples from the site can be obtained that suggest the site is dirty. The cleaner the site, the less chance of this happening. Conversely, a site may be dirty, but the particular set of samples suggest the site is clean. Again the dirtier the site, the less chance of this occurring. The chances of these errors are controlled by the position and shape of the power curve relative to the cleanup standard. Figures A.1 - A.4 illustrate several families of power curves. The ideal shape of a power curve is a step function that has a 1.0 probability of declaring the site clean whenever the true concentration is less than the cleanup standard and a zero probability of declaring the site clean when the concentration is greater than the cleanup standard.

The second use of a power curve is to help decide on an appropriate sample size for a sampling program. The lower the variability and the more samples taken, the closer the power curve will come to approaching the ideal step function described above. In addition, the trade-off between the false positive and negative rate influences the position of the power curve. Use the power curves in Appendix A to assist with the sample size determination process in one of two ways:

- Select the power curve desired for the statistical test and determine from this the sample size that is required; or
- Select the sample size to be collected and determine what the resulting power curve will be for the statistical procedure.

Chapters 6, 7, and 8 provide specific methods for making sample size determinations.

2.3 Attainment or Compliance Criteria

The characteristic of the chemical concentrations to be compared to the cleanup standard must be specified in order to define a statistical test to determine whether a sample area attains the cleanup standard. Such characteristics might be the mean concentration, the median, or the 95th percentile of the concentrations. In other words, it must be decided whether the cleanup standard is intended to be applied as a mean value such that the mean of the site must be below the cleanup standard or whether the cleanup standard is a high percentile value that must rarely be exceeded at only 5 or 10 percent of the site. Figure 2.4 illustrates these characteristics on three distributions. Section 3.5 offers a more detailed discussion of these parameters.

2.3.1 Mean

The location or general magnitude of a set of data is often characterized by the mean of the distribution. The mean of the concentration distribution is the value that corresponds to the "center" of the distribution in the sense of the "center of gravity." In determining the mean from a highly skewed lognormal distribution, small amounts of soil with concentrations far above the mean are balanced by large amounts of soil with concentrations close to, but below, the mean.

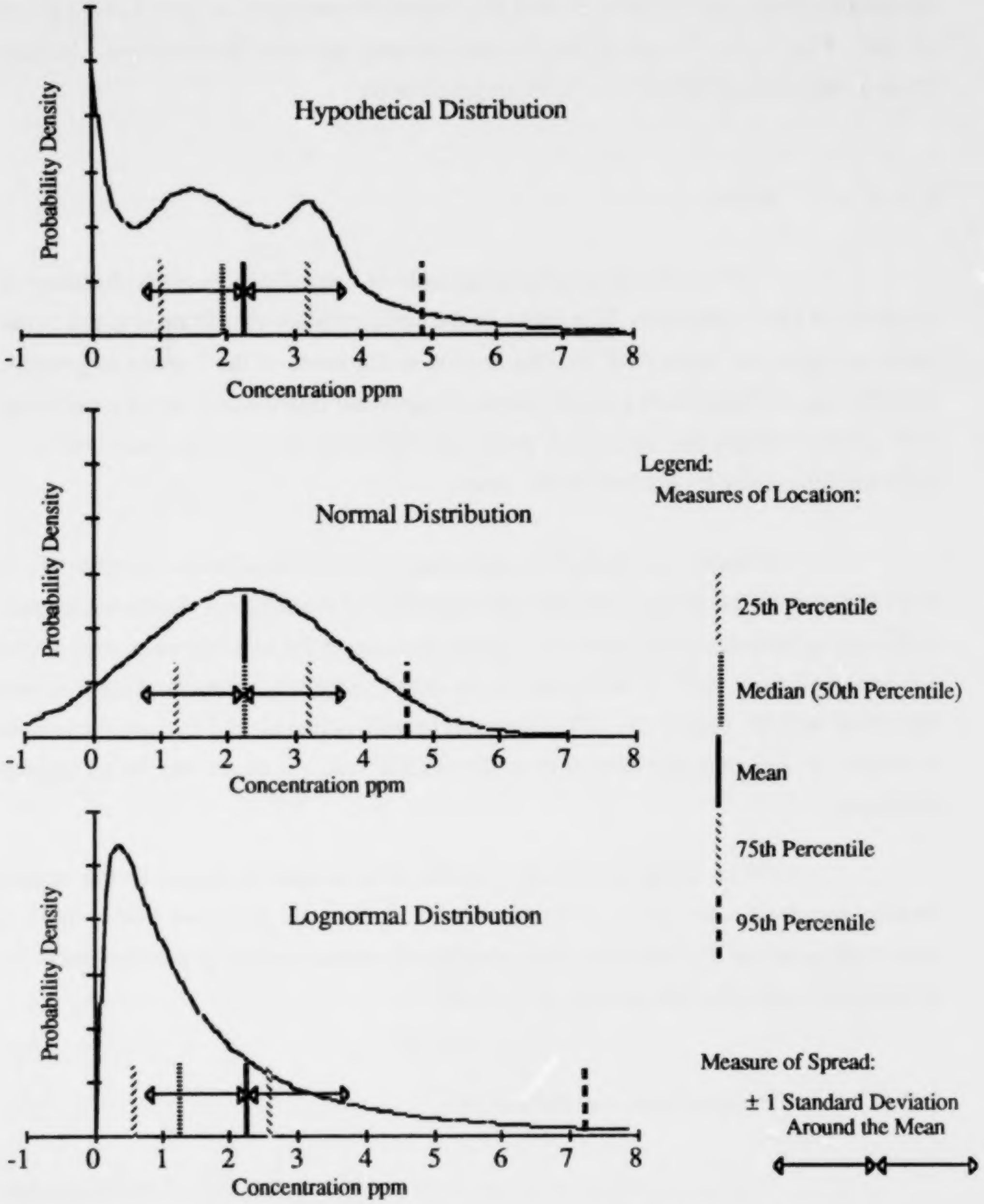
Whether the mean is a useful summary of the distribution depends on the characteristics of the sample area and the objectives of the cleanup. In a sample area with uniform contamination and very little spread or range in the concentration measurements, the mean will work well. If the spread in the data is large relative to the mean, the average conditions will not adequately reflect the most heavily contaminated parts of the population. If interest is in the average exposure or the chronic risk, the mean may be an appropriate parameter.

When using the mean, consideration should be given to the number of measurements that are likely to be recorded as below the detection limit. With many observations below the detection limit, the simple estimate of the population mean cannot be calculated (see the discussion in section 2.5.2).

2.3.2 Proportions or Percentiles

High percentiles or proportions pertain to the tail of a distribution and control against having large concentration values. The 50th percentile, the median, is often a useful alternative to the mean.

Figure 2.4 Measures of Location: Mean, Median, 25th Percentile, 75th Percentile, and 95th Percentile for Three Distributions



Methods are available for proportions that are unaffected by concentrations below the detection limit, as long as the detection limit is below the cleanup standard. The likelihood of having many data values below the detection limit makes the proportion of soil units above the cleanup standard an appealing parameter to use in assessing attainment. If the cleanup standard is only slightly above the detection limit, then it will always be possible to calculate the proportion of soil samples above the cleanup standard.

Knowing the maximum concentration of the hazardous contaminant at a waste site would be helpful in making decisions. Unfortunately, in realistic situations the maximum cannot be determined from a sample of data. A test of proportions, using an upper percentile of the concentration distribution, can serve as a reasonable approximation of the maximum value.

2.4 Components of a Risk-Based Standard

Chapter 1 introduced the concept of a risk-based standard and its application to Superfund activities. Here we will describe how statistical sampling and analysis methods can be used to adjust the stringency of a risk-based standard.

A hypothetical example of a risk-based standard is as follows: a soil concentration of arsenic greater than 20 ug/kg at a specific smelter subjects workers to a 1 in a million chance of oral cancer during a lifetime. It is commonly thought that the only way to change the stringency of the 20 ug/kg standard is to change the magnitude of the number, 20. In other words, a less stringent standard is obtained by changing the risk-based standard to 25 ug/kg with an associated increase in the probability of acquiring oral cancer. This is true, but there are other ways to influence the stringency of the standard.

There are three components of a risk-based standard that can be used to adjust its stringency. Bisgaard and Hunter (1986) provide discussion of these components and their application. The three components are:

1) The magnitude of the Concentration Threshold Level (Cs);

2) The method for obtaining data or the Sampling Plan; and

3) The evaluation scheme, Decision Rule, that will be used to compare the data with the threshold level.

Figure 2.5 illustrates the relationship among these components. The choice of a numerical level is one element of a risk standard. Other questions must also be answered regardingsampling: How many samples? In what area will the samples be obtained? In what pattern will the samples be chosen? In addition, after the data are obtained a decision framework must be developed to analyze the data. Will no more than one exceedance in 10 samples be permitted or will no more than 10 exceedances in 100 be allowed? That is, what level of confidence is required to conclude that the site is clean? Answers to these questions influence the spread of the distribution in Figure 2.1 in D, E, F, and G and, therefore, the steepness of the curve used for the Decision Rule in Figure 2.5, which is a power curve similar to Figure 2.2.

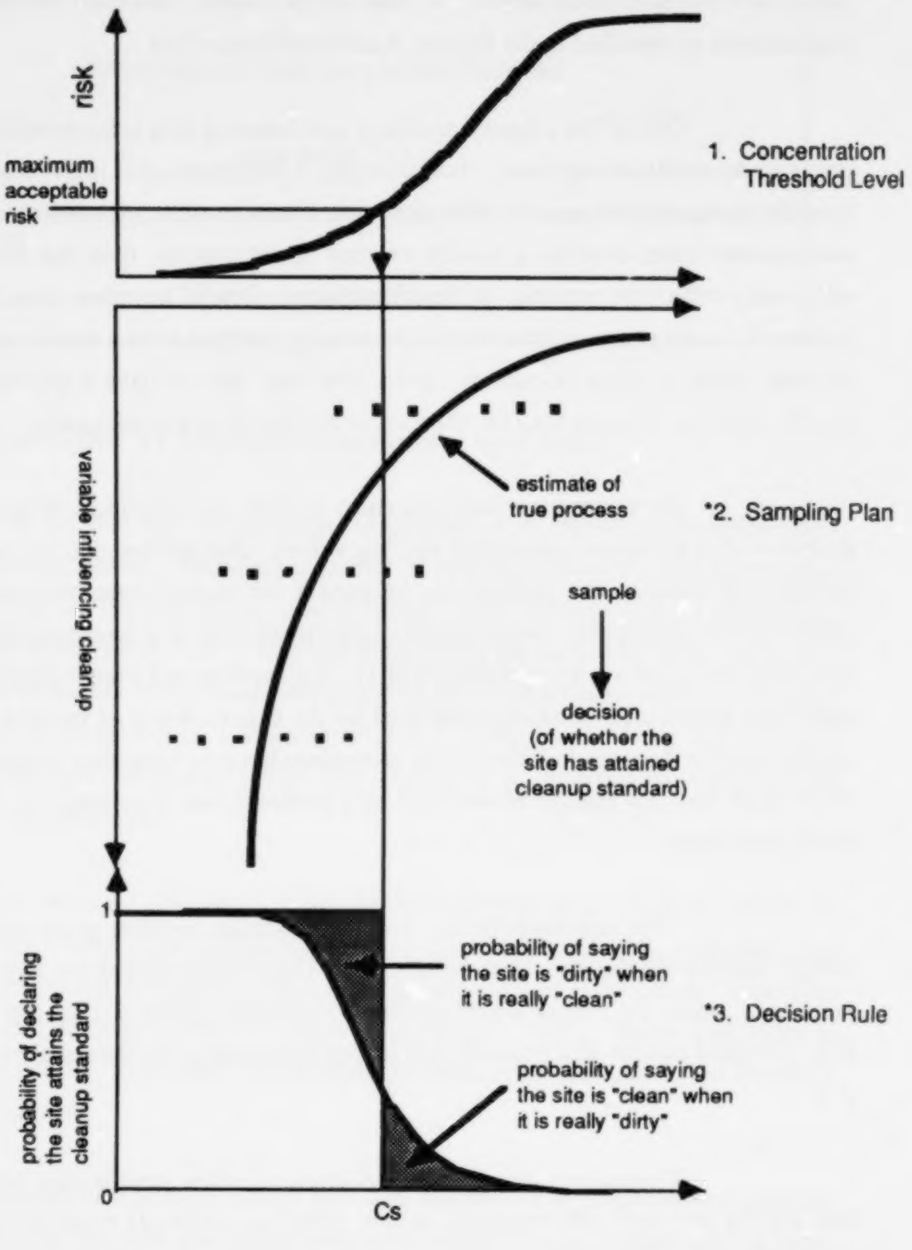
The following scenario describes the impact that the sampling plan and decision rule can have on the actual degree of cleanup. A stringent chemical concentration level is imposed as a requirement at a site (component 1). In contrast, five samples will be obtained after remediation to verify attainment of the standard (component 2), and 80 percent confidence that the new site mean is less than the standard will be required (component 3). The health effect results obtained by imposing a stringent numerical level standard are weakened because the area has not been thoroughly sampled and the associated level of confidence in the conclusions is relatively low. In this case, a poor sampling plan and low required level of confidence have influenced the actual degree of cleanup in spite of the stringency of the numerical standard.

2.5 Missing or Unusable Data, Detection Limits, Outliers

2.5.1 Missing or Unusable Data

In any sampling program, physical samples will be obtained in the field and then, some time during processing, a problem develops and a reliable measurement is not available. Samples can be lost, be labeled incorrectly, exceed holding times, be transcribed

Figure 2.5 Components of a Risk-Based Standard



risk
maximum acceptable risk
1. Concentration Threshold Level
variable influencing cleanup
estimate of true process
*2. Sampling Plan
sample
decision (of whether the site has attained cleanup standard)
1
0
probability of declaring the site attains the cleanup standard
probability of saying the site is "dirty" when it is really "clean"
*3. Decision Rule
probability of saying the site is "clean" when it is really "dirty"
Cs

incorrectly, or not satisfy quality control specifications. Clearly, missing data are not available and cannot be used in data analysis. Data that do not satisfy the most rigorous quality control specifications may or may not be usable; however, this depends on the requirements as specified in the Quality Assurance Project Plan.

One of the primary problems with missing data is the possibility that bias is imposed on statistical estimates. For example, if the presence of high concentrations of a specific contaminant causes laboratory interferences that prevent samples with the contaminant from satisfying quality control specifications, then the data set will not adequately reflect the presence of the contaminant. Careful attention should be paid to the pattern of missing data to determine if the missing samples have a similar attribute such as location, time, or chain of custody. If so, then they may all have a special concentration profile, and their absence may be affecting or biasing the result summary.

However, the main question is how can planning help to prevent the problem of an excessive number of missing values. One method can be used to help plan for missing values. The method can be used if the approximate proportion of missing values can be anticipated, based on prior experience with or a professional judgment of a sampling team, laboratory, and data analyst. The number of samples needed to conduct a particular statistical evaluation is inflated by the expected rate of missing values. More sample results than needed will not be a problem because precision will increase; on the other hand, too few sample results will be a problem, and may result in more treatment being required.

The equation for the simplest situation requires prior estimation of the sample size for the statistical procedures (n_d). This is discussed above and throughout the document. Also, the rate at which missing or unusable values occur must be determined (R). The final sample size required (n_f) is then estimated using the simple equation in Box 2.1.

Throughout this guidance document, when sample size formulae, tables, and graphs are used, the resulting sample sizes (n_d and n_{hd}) required for a statistical analysis having a specified precision can be increased using these equations in anticipation of missing data.

Box 2.1
Estimating the Final Sample Size Required

$$n_f = n_d/(1 - R)$$

A similar equation is used for each of the h strata in a stratified sampling plan:

$$n_{hf} = n_{hd}/(1 - R_h)$$

2.5.2 Evaluation of Less-Than-Detection-Limit Data

The science and terminology associated with less-than-detection-limit chemistry are unstandardized. There are a variety of opinions, methods, and approaches for reporting chemicals present at low concentration. The problem can be segmented. First, there is the problem of how a chemist determines the detection limit value and EXACTLY what it means when values are reported above and below a detection limit. This question is not the subject of this document, but it is important. There is substantial literature on this subject and Bishop (1985) and Clayton et al. (1986) offer useful insight and access to other references.

The second problem is: How should less-than-detection-limit values be evaluated along with other values larger than the detection limit when both are present in a data set? This subject also is supported by a considerable amount of literature. Examples include Gilbert and Kinnison (1981); Gilliom and Helsel (1986); Helsel and Gilliom (1986); and Gleit (1985). This aspect of the detection limit problem is discussed briefly in the following paragraph.

Fortunately, because of the null and alternative hypothesis arrangement, having concentrations less then a detection limit is no problem when a proportion is being tested, provided the detection limit is less than Cs. When the proportion or percentile is being tested, the important attribute of each data value is whether it is larger or smaller than the Cs, rather than the magnitude of the value. In fact, a site can be evaluated easily relative to a high percentile in spite of a data set that includes many values less than the detection limit, which is expected when a cleanup technology has uniformly reduced most concentration measurements to less than the detection limit.

When the mean is being used as the basis of comparison with a cleanup standard, the magnitude of each value is important. When values are reported as being less than a detection limit, it is generally recommended that they be included in the analysis as values at the detection limit. This method accommodates detection limits that vary across samples, and the method is simple to use. In addition, this approach, although statistically biased, errs in favor of health and environmental protection because of the construction of the null and alternative hypothesis described earlier. In some cases a less-than-detection-limit value may be quite large relative to other measured values in a data set. In this case it may be best to delete such a value. Other methods are available for statistically addressing less-than-detection-limit values as described above, but they may not be as conservative with respect to environmental protection.

2.5.3 Outliers

Measurements that are extremely large or small relative to the rest of the data gathered and that are suspected of misrepresenting the true concentration at the sample location are often called "outliers." If a particular observation is suspected to be in error, the error should be identified and corrected, and the corrected value used in the analysis. If no such verification is possible, a statistician should be consulted to provide modifications to the statistical analysis that account for the suspected "outliers." Methods to detect and accommodate outliers are described in Barnett and Lewis (1978) and Grubbs (1969).

The handling of outliers is a controversial topic. This document recommends that **all data not known to be in error should be considered valid** because:

- The expected distribution of concentration values may be skewed (i.e., nonsymmetric) so that large concentrations, which look like "outliers" to some analysts, may be legitimate;

- The procedures recommended in this document are less sensitive to extremely low concentrations than to extremely high concentrations; and
- High concentrations are of particular concern for their potential health and environmental impact.

2.6 General Assumptions

The statistical procedures recommended in this guidance document must be applicable to many different field situations; therefore, the procedures that have been chosen are generally based on few assumptions. Situations in which other statistical procedures might be used to provide more accurate or more cost-effective results will be noted with references.

This document assumes that (1) the sources of contamination and contaminating chemicals are known, (2) the sources of contamination have been removed, or there is no reason to believe that the concentrations of contaminant in the soil will increase after treatment, and (3) chemical concentrations do not exhibit short-term variability over the sampling period. The methods presented can be used if sources of contamination exist or concentrations are expected to increase. However, sampling may have to be repeated and the results carefully interpreted and presented to reflect the possibility of additional contamination.

When statistical tests are repeated to evaluate several chemicals, such as testing that concentration levels for two chemicals both attain the cleanup standard, it is assumed that the sample area will be declared to attain the cleanup standard only if all statistical tests used are consistent with this conclusion. For other procedures that might be used to combine the results of individual tests, it would be advantageous to consult a statistician.

2.7 A Note on Statistical Versus Field Sampling Terminology

The term **sample** is used in two different ways. One refers to a physical soil sample collected for laboratory analysis, and the other refers to a collection of data called a statistical sample. To avoid confusion, definitions of several terms follow.

Physical sample or soil sample: A portion of material (such as a soil core, scoop, etc.) gathered at the waste site on which laboratory measurements are to be made. This may also be called a **soil unit**.

Statistical sample: A statistical sample consists of the collection of multiple physical samples obtained for assessing attainment of the cleanup standard. The units included in a statistical sample are selected by probabilistic means.

Sample: The word "sample" in this manual will generally have the meaning of "statistical sample."

Sample size: The number of soil units being measured or the size of the statistical sample. Thus, a sample of size 10 consists of the measurements taken on 10 soil units.

Size of the physical sample: This term refers to the volume or weight of a soil unit or the quantity of soil in a single physical sample.

The following terms refer to the manner in which the statistical sample of physical samples is collected: **random sample, systematic sample, stratified sample, judgment sample**. These sample designs are discussed in Chapter 4.

2.8 Summary

Errors are possible in evaluating whether or not a site attains the cleanup standard. For example, consider the errors associated with an extreme decision rule where no exceedances of a standard are allowed. The site may be dirty even when substantial sampling shows no exceedances; however, one sample may exceed the cleanup standard and the site is judged dirty even when the site is acceptably clean.

Statistical methods provide approaches for balancing these two decision errors and allow extrapolation in a scientifically valid fashion. This chapter reviews the statistical concepts that are assumed and used as part of the procedures described in this guidance document. These include:

- A false positive decision--that the site is thought to be clean when it is not;
- A false negative decision--that the site is thought to be contaminated when it is not;
- The factors that specify the shape and location of the power curve relative to the cleanup standard and to sample size determination;
- The mean--the value that corresponds to the "center" of the concentration distribution;
- Proportions or percentiles--a value that can be used effectively, based on the distribution of contaminant concentration, to approximate the maximum concentration of the hazardous contaminant.

The components of a risk-based standard and how these components relate to one another are reviewed and graphically illustrated. Methods to help plan for missing or unusable data, less-than-detection-limit data, and outliers are discussed, followed by the general assumptions associated with the statistical procedures explained in this document. These assumptions are that:

- All of the sources of contamination and contaminating chemicals are known;
- These sources have been removed, so that the contamination will not increase after treatment; and
- Chemical concentrations do not exhibit short-term variability over the sampling period.

3. SPECIFICATION OF ATTAINMENT OBJECTIVES

The specification of attainment objectives must be completed by personnel familiar with:

- The engineering aspects of the remediation;
- The nature and extent of contamination present;
- Health and environmental risks of the chemicals involved; and
- The costs of sampling, analysis, and cleanup.

Attainment objectives are the procedures and criteria that must be defined to guide waste site managers and personnel in the process of sampling and data analysis to achieve a predetermined cleanup standard. Meeting these objectives and criteria enable the waste site to be judged sufficiently remediated.

As indicated in Figure 1.1, defining attainment objectives is the first task in the evaluation of whether a site has attained a cleanup standard. Figure 3.1 divides the box devoted to the establishment and definition of cleanup objectives into its components.

3.1 Specification of Sample Areas

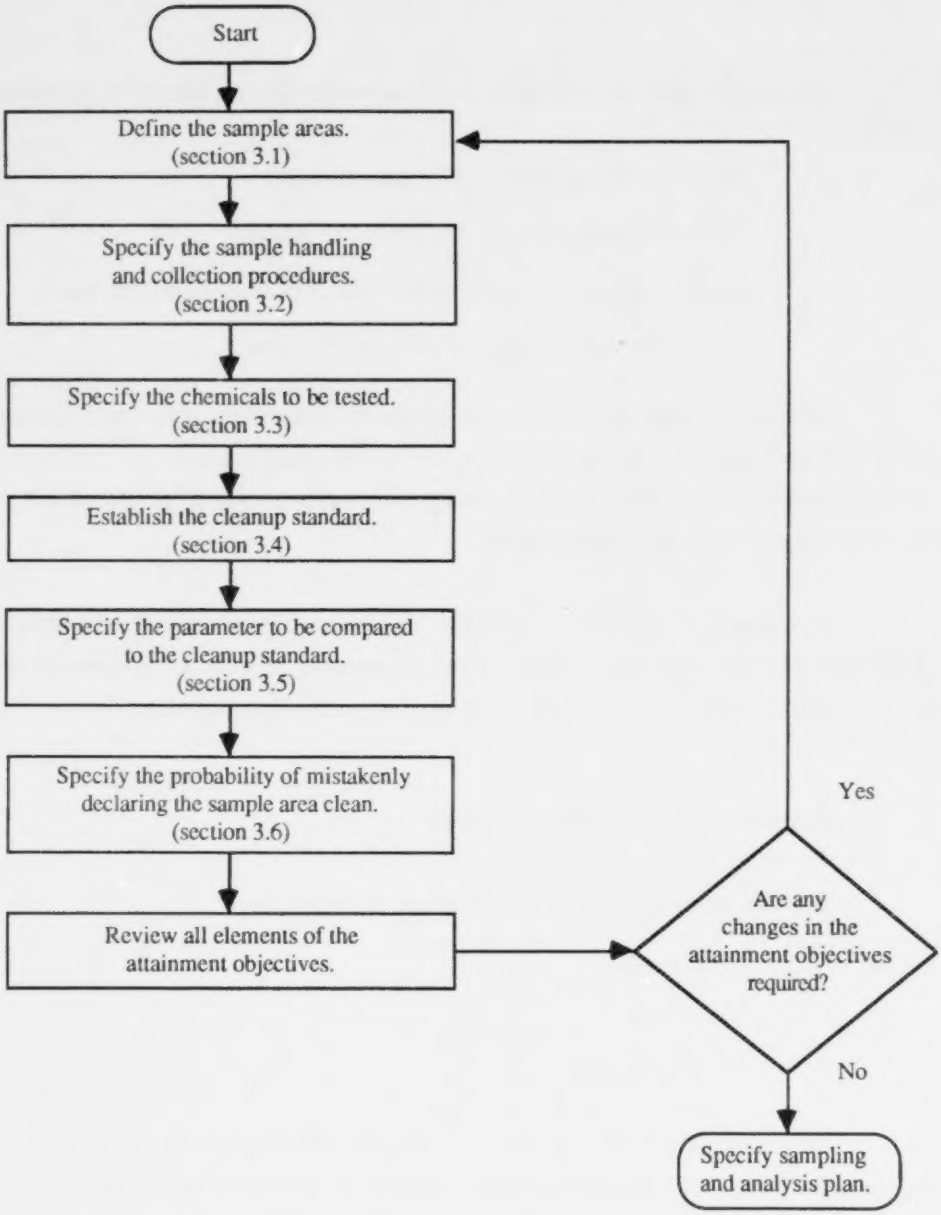
Three terms describing areas within the waste site are:

- Sample area;
- Strata; and
- Sample location.

These terms are used in establishing the attainment objectives and the sampling and analysis plans. Sample area specification is discussed below and methods for defining strata and sample locations are discussed in Chapter 5.

The waste site should be divided into sample areas. Each sample area will be evaluated separately for attainment of a cleanup standard and will require a separate statistical sample.

Figure 3.1 Steps in Defining the Attainment Objectives


Start
Define the sample areas. (section 3.1)
Specify the sample handling and collection procedures. (section 3.2)
Specify the chemicals to be tested. (section 3.3)
Establish the cleanup standard. (section 3.4)
Specify the parameter to be compared to the cleanup standard. (section 3.5)
Specify the probability of mistakenly declaring the sample area clean. (section 3.6)
Review all elements of the attainment objectives.
Are any changes in the attainment objectives required?
Yes
No
Specify sampling and analysis plan.

Consider the following example, which emphasizes the importance of the sample area definition. A site consists of an open field with little contamination and a waste pile covering one-quarter of the site. If sampling and data analysis were executed without respect to the waste pile, it might be maintained that the mean concentration of the site was statistically lower than the standard. The site wide mean was "excessively" low because the waste pile data were "diluted" by many open field measurements. The solution is to define the waste pile as one sample area and the open field as another. Attainment decisions will be made independently for each area.

Because of the potential for this problem, it is important to ensure that sample areas are clearly defined during the design phase. Parties must agree that if the sample area is judged clean, no more cleanup is required in any part of the sample area. There are several considerations associated with the definition of sample areas.

1) It is generally useful to define multiple sample areas within a waste site. These areas should be defined so that they are as homogeneous as possible with respect to prior waste management activities. For example, if a PCB transformer disposal area and a lead battery recycling area are located on the same site, they should not be included in the same sample area.

2) It may also be useful to define sample areas by batches of material that will receive a treatment action, for example, dump truck loads (see Exner et al., 1985) or the minimum sized areas that can be stabilized or capped.

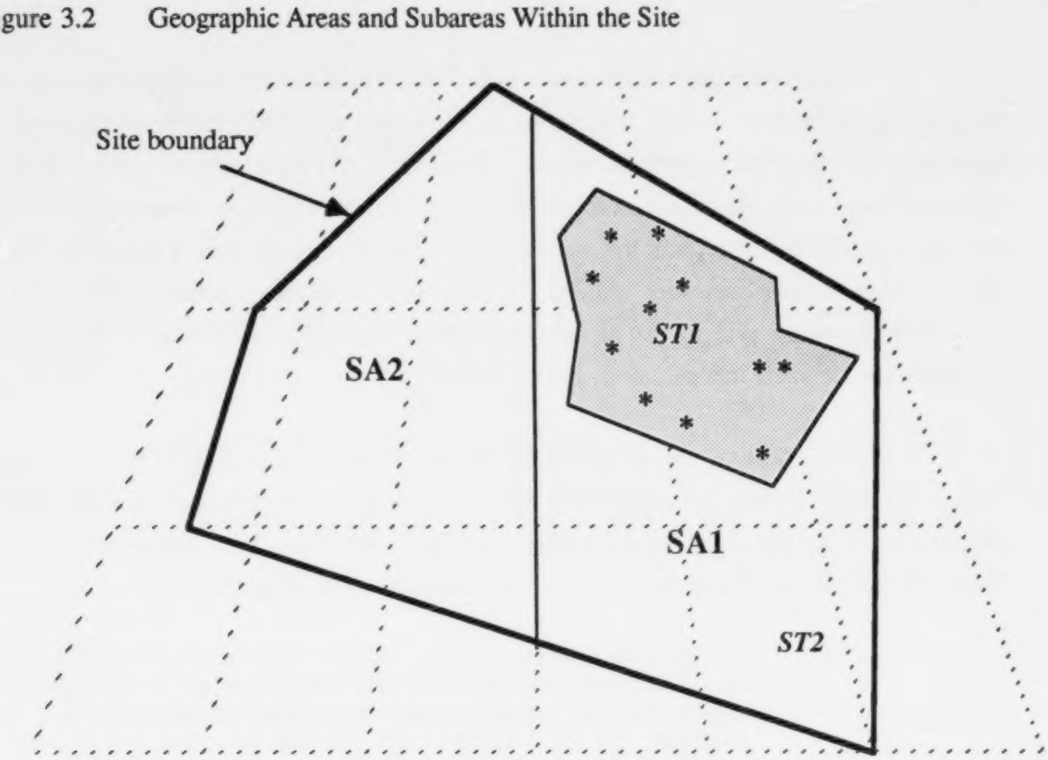
3) A site may be comprised of areas that require different sampling or treatment technologies. For example, disturbed versus natural soils, wetlands versus firm terrain, or sandy versus clay soils may suggest establishment of different sample areas.

4) Finally, while more (smaller) sample areas provide more flexible response to changing conditions, sampling costs will increase with the number of sample areas.

5) Sample area definitions also require that the depth or depth intervals of interest be specified. This is discussed in greater detail in section 5.6.

Figure 3.2 shows how different geographic sample areas relate to one another.

Figure 3.2 Geographic Areas and Subareas Within the Site



Map of the waste site

Waste site with two sample areas, **SA1** and **SA2**. Separate attainment decisions are made for each sample area. **Sample area SA1** is divided into two strata, ***ST1*** and ***ST2.*** (See Chapter 4 for more on stratified sampling [multiple strata per sample area].) **Stratum** ***ST1*** has randomly selected soil sample locations indicated by "*".

3.2 Specification of Sample Collection and Handling Procedures

Deciding whether a sample area attains the cleanup standard requires that measurements be made on a statistical sample of soil units, and that these measurements be compared to the cleanup standard. An important task for any decision procedure is to define carefully what is being measured; questions that must be answered include:

- What is meant by a soil unit or soil sample?

- How is the soil sample collected and what equipment and procedures are used?
- How is the soil sample handled between collection and measurement?
- How are the laboratory measurements to be made and what accuracy is to be achieved?

The above questions are not addressed in this document. Consult the guidances listed in Table 1.1 for more information.

3.3 Specification of the Chemicals to be Tested

For each sample area, the chemicals to be tested in each soil unit should be listed. When multiple chemicals are tested, this document assumes that all chemicals must attain the cleanup standard for the sample area to be declared clean.

3.4 Specification of the Cleanup Standard

Concentration measurements for each physical sample will be compared to the appropriate, relevant, or applicable cleanup standard chosen for each chemical to be tested. Cleanup standards are determined by EPA during the site-specific endangerment assessments. The cleanup standard for each chemical of concern must be stated at the outset of the remedial verification investigation. Final selection of the cleanup standard depends on many factors as discussed in USEPA (1986c). Selection of the cleanup standard depends on the following factors:

- The availability and value of other appropriate criteria;
- Factors related to toxicology and exposure, for example, the effect of multiple contaminants, potential use of the waste site and pathways of exposure, population sensitivities to the chemical;
- Factors related to uncertainty, for example, the effectiveness of treatment alternatives, reliability of exposure data, and the reliability of institutional controls; and
- Factors related to technical limitations, for example, laboratory detection limits, background contamination levels, and technical limitations to restoration.

Throughout this document, the cleanup standard will be denoted by Cs.

3.5 Selection of the Statistical Parameter to Compare with the Cleanup Standard

3.5.1 Selection Criteria for the Mean, Median, and Upper Percentile

Criteria for selecting the parameter to use in the statistical assessment decision are:

- The criteria used to develop the risk-based standards, if known;
- The toxicological effect of the contaminant being measured (e.g., carcinogenic, systemic toxicant, developmental toxicant).
- The relative sample sizes required or the relative ease of calculation;
- The likelihood of concentration measurements below the cleanup standard; and
- The relative spread of the data.

Table 3.1 presents these criteria and when they support or contradict the use of the mean, upper percentile, and median. The median may offer a reasonable compromise because the median is the 50th percentile and a measure of central tendency. Table 3.2 illustrates the broad potential utility of the median.

Table 3.1 Points to consider when trying to choose among the mean, high percentile, or median

Parameter	Points to Consider
Mean	1) Easy to calculate and estimate a confidence interval. 2) Requires fewer samples than other parameters to achieve similar confidence. 3) Useful when the cleanup standard has been based on consideration of carcinogenic or chronic health effects or long-term average exposure. 4) Useful when the soil is uniform with little spread in the sample data. 5) Not as useful when contamination exists in small areas within a larger area that is being sampled because the mean can be "diluted" or reduced by the inclusion of clean areas in the sample area. 6) Not very representative of highly variable soils because the most heavily contaminated areas are not characterized by a mean. 7) Not useful when there are a large proportion of less-than-detection-limit values.
Upper Proportion/ Percentile	1) Can be expressed in terms that have more meaning than tests of the mean. Volumes or areas can be expressed relative to the total volume or area of concern, and this can be a proportion of importance. For example, if no more than 10,000 m^3 in a total volume of 1,000,000 m^3 can exceed a cleanup standard, then this becomes a test to verify with reasonable confidence that no less than 99 percent of the site is below the cleanup standard. 2) Will provide the best control of extreme values when data are highly variable. 3) Some methods are unaffected by less-than-detection-limit values, as long as the detection limit is less than the cleanup standard. 4) If the health effects of the contaminant are acute or worst-case effects, extreme concentrations are of concern and are best evaluated by ensuring that a large proportion of the site is below a cleanup standard.

Table 3.1 Points to consider when trying to choose among the mean, high percentile, or median (continued)

Parameter	Points to Consider
Upper Proportion/ Percentile (continued)	5) Similar to the mean, if contamination exists within a small area, but if the sampling program is conducted to include a much larger surrounding area with little contamination, the proportion will be affected or "diluted." 6) The proportion of the site that must be below the cleanup standard must be chosen. 7) When statistical methods are used that require few assumptions, a larger sample size will be required than for tests based on the mean.
Median	1) Has benefits over the mean because it is not as heavily influenced by outliers and highly variable data, and can be used with a large number of less-than-detection-limit values. 2) Has many of the positive features of the mean, in particular its usefulness for evaluating cleanup standards based on carcinogenic or chronic health effects and long-term average exposure. 3) Has positive features of the proportion, including its reliance on fewer assumptions. 4) Retains some negative features of the mean in that the median will not control extreme values.

Table 3.2 Recommended parameters to test when comparing the cleanup standard to the average concentration of a chemical with chronic effects

Data Variability	Proportion of the data with concentrations below the detection limit: Low (Perhaps < 50%)	High (Perhaps > 50%)
Large Coefficient of Variation (Perhaps cv > .5)	Mean (or Median)	Upper Percentile
Small Coefficient of Variation (Perhaps cv < .5)	Mean (or Median)	Median

3.5.2 Multiple Attainment Criteria

This guidance document addresses testing for a single parameter--the mean or a specified percentile of the distribution--that is below the cleanup standard. However, in some situations two or more parameters can be chosen. The sample area would be declared clean if all parameters were significantly less than the cleanup standard. For example, there may be interest in providing protection against excessive extreme and average concentrations. Therefore, the mean and an upper percentile can be tested using the rule that the sample area attains the cleanup standard if **both** parameters are below the cleanup standard. When testing both parameters, the number of samples collected will be either the number required for the test of the mean or the number required for the test of the percentile (whichever number is larger).

Other more complicated criteria may be used to assess the attainment of the cleanup criteria. Multiple criteria are established in the following examples. In each case it is desirable that:

- Most of the soil has concentrations below the cleanup standard and that the concentrations above the cleanup standard are not too large.

This may be accomplished by testing whether the 75th percentile is below the cleanup standard and whether the mean of those concentrations above the cleanup standard is less than twice the cleanup standard. This combination of tests can be performed with minor modifications to the methods presented in this document.

- The mean concentration be less than the cleanup standard and that the standard deviation of the data be small, thus limiting the number of extreme concentrations. This may be accomplished by testing if the mean is below the cleanup standard and the coefficient of variation is below some low level (.5 for example). This document does not address testing the standard deviation, variance, or coefficient of variation against a cleanup standard.
- The mean concentration be less than the cleanup standard and that the remaining contamination be uniformly distributed across the sample area relative to the overall spread of the data. Testing these criteria may be accomplished by testing for a mean below the cleanup standard and variability between strata means that is not large compared to the variability within strata (analysis of variance).
- The mean concentration be less than the cleanup standard and that no area of contaminated soil (assumed to be circular) be larger than a specified size. Testing these criteria involves testing for hot spots, which are discussed in Chapter 9 and more extensively in Gilbert (1987).

3.6 Decision Making With Uncertainty: The Chance of Concluding the Site Is Protective of Public Health and the Environment When It Is Actually Not Protective

As discussed in Chapter 2, the validity of the decision that a site meets the cleanup standard depends on how well the samples of soil represent the site, how accurately the soil samples are analyzed, and other factors, all of which are subject to variation. Different sampling patterns will yield different results and repeated measurements on individual soil samples will yield different concentrations. This variation introduces uncertainty into the decision concerning the attainment of a cleanup standard.

As a result of this uncertainty, one may decide that the site is clean when it is not. In the context of this document, this mistaken conclusion can be referred to as a **false positive** finding (the chance or probability of a false positive is indicated by the Greek letter alpha, α). There are two important points surrounding false positives:

- First, from an environmental and health protection perspective, it is imperative to reduce the chance of a false positive. In direct terms a false positive is the chance of deciding a Superfund site is clean when it still poses a health or environmental threat. Of course, a low false positive rate does not come without a cost. The additional cost required to lower the false positive rate comes from additional samples and more accurate sampling and analysis methods.
- Second, the definition of a false positive in this document is exactly opposite the more familiar definition of a false positive under RCRA detection and compliance monitoring. This is because the null and alternative hypotheses are reversed, once a site has been verified to have contamination. Under the RCRA detection monitoring situation, EPA was concerned about a high false negative rate; here EPA is concerned about a high false positive rate.

In order to design a statistical test for deciding whether the sample area attains the cleanup standard, those individuals specifying the sampling and analysis objectives should select and specify the false positive rate for testing the site. While different false positive rates can be used for each chemical, it is recommended that all chemicals in the sample area use the same rates. This rate is the maximum probability that the sample area will be declared clean by mistake when it is actually dirty. For a further discussion of false positive rates, see Sokal and Rohlf (1981).

3.7 Data Quality Objectives

The Quality Assurance Management staff within EPA has developed requirements and procedures for establishing Data Quality Objectives (DQOs) when environmental data are collected to support regulatory and programmatic decisions. The DQOs are a clear set of statements addressing the following issues (see USEPA, 1987a and USEPA, 1987b).

- The decision to be made;
- The reasons environmental data are needed and how they will be used;
- Time and resource constraints on data collection;
- Detailed description of the data to be collected;

- Specifications regarding the domain of the decision;
- The consequences of an incorrect decision attributable to inadequate environmental data;
- The calculations, statistical or otherwise, that will be performed on the data in order to arrive at the result, including the statistic that will be used to summarize the data and the "action level" (cleanup standard) to which the summary statistic will be compared; and
- The level of uncertainty that the decision maker is willing to accept in the results derived from the environmental data.

The specification of attainment objectives that have been discussed in this chapter and the sampling and analysis plan discussed in the next chapter are an important part of the Data Quality Objectives process. Completion of the DQO process will provide the required information for the specification of attainment objectives.

3.8 Summary

The following steps must be taken to evaluate whether a site has attained the cleanup standard:

- Define the attainment objectives;
- Specify sample design and analysis plan, and determine sample size;
- Collect the data; and
- Determine if the sample area attains the cleanup standard.

This chapter discusses attainment objective specifications. Attainment objectives are specified by RPMs, RPs, and their contractors. They are not statistically based decisions.

- **Define the sample area.** The waste site should be divided into sample areas. Each sample area will be evaluated separately for attainment of a cleanup standard and will require a separate statistical sample. It is important to ensure that sample areas are clearly defined during the design phase.
- **Specify the sample handling and collection procedures.** An important task for any decision procedure is to define carefully what is being measured.

- **Specify the chemicals to be tested.** Chemicals to be tested in each soil unit should be listed.

- **Establish the cleanup standard.** Cleanup standards are determined by EPA using site-specific risk assessments or ARARs. The cleanup standard for each chemical of concern must be stated at the outset of the remedial verification investigation.

- **Specify the parameter to be compared to the cleanup standard.** In other words: "Does the cleanup standard represent an average condition (mean) or a level to be rarely exceeded (high percentile)? Criteria for selecting the parameter to use in the statistical assessment decision are:

    - The criteria used to develop the risk-based standards, if known;

    - Whether the contaminant being measured has an acute or long-term chronic effect;

    - The relative sample sizes required or the relative ease of calculation;

    - The likelihood of concentration measurements below the detection limit; and

    - The relative spread of the data.

- **Specify the probability of mistakenly declaring the sample area clean.** Select and specify the false positive rate for testing the site. It is recommended that all chemicals in the sample area use the same rates. This rate is the maximum probability that the sample area will be declared clean by mistake when it is actually dirty.

- **Review all elements of the attainment objectives.**

4. DESIGN OF THE SAMPLING AND ANALYSIS PLAN

Once the attainment objectives are specified by program and subject matter personnel, statisticians can be useful for designing important components of sampling and analysis plans.

The methods of analysis must be consistent with the sample design and the attainment objectives. For example, data that are collected using stratified sampling cannot be analyzed using the equations for simple random sampling. The sample design and analysis plan must coincide. If there appears to be any reason to use different sample designs or analysis plans than those discussed in this manual, or if there is any reason to change either the sample design or the analysis plan after field data collection has started, it is recommended that a statistician be consulted.

This chapter presents some approaches to the design of a sampling and analysis plan and presents the strengths and weaknesses of various designs.

4.1 The Sampling Plan

The following sections provide background discussion guiding the choice of sampling plan for each sampling area. Chapter 5 discusses the details of how to implement a sampling plan. For more details, see Kish (1965), Cochran (1977), Hansen et al. (1953), or the EPA guidances in Table 1.1.

The sample designs considered in this document are:

- Simple random sampling called **random sampling** in this document;
- Stratified random sampling called **stratified sampling** in this document;
- Simple systematic sampling called **systematic sampling** in this document; and

- Sequential random sampling called **sequential sampling** in this document.

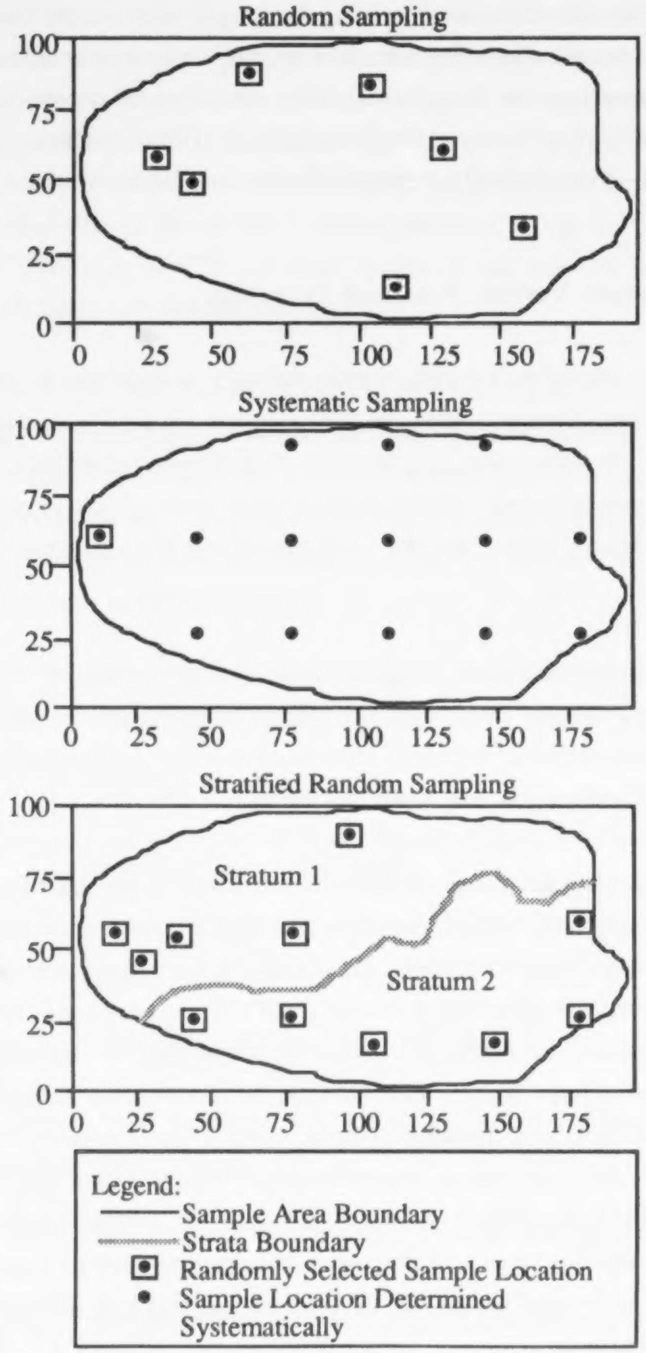
Randomization is necessary to make probability or confidence statements about the results of the sampling. Both random and random start systematic sample locations have random components. In contrast, sample selection using the judgment of the sampler has no randomization. Results from such samples cannot be generalized to the whole sample area and no probability statements can be made when judgment sampling is used. Judgment sampling may be justified, for example, during the preliminary assessment and site investigation stages if the sampler has substantial knowledge of the sources and history of contamination. However, judgment samples should not be used to determine whether the cleanup standard has been attained.

Combinations of the designs referred to above can also be used. For example, systematic sampling could be used with stratified sampling. In the situation where cleanup has occurred, if the concentrations across the site are relatively low and uniform and the site is accessible, the sample designs considered in this document should be adequate. If other more complicated sample designs are necessary, it is recommended that a statistician be consulted on the best design, and on the appropriate analysis method for that design. Figure 4.1 illustrates a random, systematic, and stratified sample.

4.1.1 Random Versus Systematic Sampling

Random selection of sample points requires that each sample point be selected independent of the location of all other sample points. Figure 4.1 shows a random sample. Note that under random sampling no pattern is expected in the distribution of the points. However, it is possible (purely by chance) that all of the sample points will be clustered in, say, one or two quadrants of the site. This possibility is extremely small for larger sample sizes.

Figure 4.1 Illustration of Random, Systematic, and Stratified Sampling (axes are distance in meters)


Random Sampling
Systematic Sampling
Stratified Random Sampling
0
25
50
75
100
125
150
175
Stratum 1
Stratum 2
Legend:
Sample Area Boundary
Strata Boundary
Randomly Selected Sample Location
Sample Location Determined Systematically

An alternative to random sampling is systematic sampling, which distributes the sample more uniformly over the site. Because the sample points follow a simple pattern and are separated by a fixed distance, locating the sample points in the field may be easier using a systematic sample than using a random sample. In many circumstances, estimates from systematic sampling may be preferred. More discussion of systematic versus random sampling can be found in Finney (1948), Legg, et al. (1985), Cochran (1977), Osborne (1942), Palley and Horwitz (1961), Peshkova (1970), and Wolter (1984).

4.1.2 Simple Versus Stratified Sampling

The precision of statistical estimates may be improved by dividing a sample area into more homogeneous strata. In this way, the variability due to soil, location, characteristics of the terrain, etc. can be controlled, thereby improving the precision of contamination level estimates. Homogeneous areas from which separate samples are drawn are referred to as "strata," and the combined sample from all areas is referred to as a "stratified sample."

Like systematic sampling, stratification provides another way of minimizing the possibility that important areas of the site will not be represented in the sample. Note in Figure 4.1 that the two strata represent subareas for which representation in the sample will be guaranteed under a stratified sampling design.

The main advantage of stratification is that it can result in a more efficient allocation of resources than would be possible with a simple random sample. For example, suppose that, based on physical features, the site can be divided into a hilly and a flat area, and that the hilly area comprises about 75 percent of the total area and is more expensive to sample than the flat area. If there is no reason to analyze the two subareas separately, we might consider selecting a simple random sample of soil units across the entire site. However, with a simple random sample, about 75 percent of the sample would be in the hilly, and therefore more expensive, areas of the site. With stratified sampling, the sample can be allocated disproportionately to the two subareas, i.e., sample fewer units from hilly areas and more from flat areas. In this way, the resulting cost savings (over a simple random sample) can be used to increase the total sample size and, hence, the precision of estimates from the sample.

The above illustration is highly simplified. In addition to differential stratum costs, factors such as the relative sizes of the strata and the variability of the contaminant under study in the different strata will affect the optimum allocation. The illustration does, however, point out that stratification can be used to design a more efficient sample, and is more than simply a device to ensure that particular subareas of the site are represented in the sample. A formal discussion of stratified sampling, and the cost and variance considerations used to determine an optimum allocation, is beyond the scope of this manual. However, sections 5.4 and 6.4 offer a discussion of the basic principles used to guide the design of a stratified sample.

Although stratified sampling is more difficult to implement in the field and slightly more difficult to analyze, stratified sampling will provide benefits if differences in mean concentrations or sampling costs across the sample area exist and can be reasonably identified using available data. It is important to define strata so that the physical samples within a stratum are more similar to each other than to samples from different strata. Factors that can be used to define strata are:

- Sampling depth (see section 5.6 for details);
- Concentration level;
- Physiography/topography;
- The presence of other contaminants that affect the analytical techniques required at the lab;
- The history and sources of contamination over the site;
- Previous cleanup attempts; or
- Weathering and run-off processes.

There are two fundamental and important points to remember when defining areas that will become different strata:

- The strata must not overlap--no area within one strata can be within another strata; and
- The sum of the sizes of the strata must equal the area of the sample area.

In other words, the strata must collectively account for the entire sample area of interest--no more, no less.

4.1.3 Sequential Sampling

For most statistical methods, the analysis is performed after the entire sample has been collected and the laboratory results are complete. In sequential random sampling, the samples are analyzed as they are collected. A statistical analysis of the data, after each sample is collected and analyzed, is used to determine if another sample is to be collected or if the sampling program terminates with a decision that the site is clean or dirty. (Sequential sampling is the subject of Chapter 8.)

4.2 The Analysis Plan

Similar to sampling plan designs, planning an approach to analysis and the actual analysis begin before the first sample is collected. The first task of the analysis plan is to determine how the cleanup standard should function. In other words, what is the cleanup standard: a value that should be rarely exceeded; an average value; or a level that defines the presence of a hot spot? This must be decided because it determines what analysis method will be used to determine attainment.

Second, the analysis plan must be developed in conjunction with the sampling plan discussed earlier in this chapter. For example, plans to conduct stratified sampling cannot be analyzed using the equations for random sampling.

Third, the first actual step required in the analysis plan should be a determination of the appropriate sample size. This requires calculations and evaluation before the data are collected. Often the number of samples is determined by economics and budget rather than an evaluation of the required accuracy. Nevertheless, it is important to evaluate the accuracy associated with a prespecified number of samples.

Fourth, the analysis plan will describe the evaluation of the resulting data. Chapters 6 through 10 offer various analytical approaches, depending on attainment objectives and the sampling program. Table 4.1 presents where in this document various combinations of analysis and sampling plans are discussed.

Table 4.1 Where sample designs and analysis methods for soil sampling are discussed in this document

Type of Evaluation	Analysis Method	Chapter Location Sample Design Random	Stratified	Systematic	Sequential
Test of the Mean	Test for means	6.3.3	6.4.2	6.5.2	
Test of Percentiles	Nonparametric	7.3.3	7.5.2	7.6	
	Tolerance Intervals	7.3.6			
	Sequential Sampling				8.2
Hot Spot Evaluation				9.2.1	
Geostatistics	Indicator Kriging			10.3	

4.3 Summary

Design of the sampling and analysis plan requires specification of attainment objectives by program and subject matter personnel. The sampling and analysis objectives can be refined with the assistance of statistical expertise. The sample design and analysis plans go together; therefore, the following methods of analysis must be consistent with the sample design:

- Random sampling;
- Stratified sampling;

- Systematic sampling; and
- Sequential sampling.

Random selection of sample points requires that each sample point be selected independent of the location of all other sample points. An alternative to random sampling is systematic sampling, which distributes the sample more uniformly over the site. Systematic sampling is preferred in hot spot searches and in geostatistical studies.

Like systematic sampling, stratified sampling minimizes the possibility that important areas of the site will not be represented by dividing a sample area into homogeneous subareas. The main advantage of stratification is that it can result in a more efficient allocation of resources than would be possible with a random sample.

Sequential sampling (Chapter 8) requires that the samples be analyzed as they are collected.

Decisions required to plan an approach to analysis are:

- Determine the analysis method that is most useful;
- Develop the plan in conjunction with the sampling plan;
- Determine the appropriate sample size; and
- Describe how the resulting data will be evaluated.

5. FIELD SAMPLING PROCEDURES

The procedures discussed in this chapter ensure that:

- The method of establishing soil sample locations in the field is consistent with the planned sample design;
- Each sample location is selected in a nonjudgmental and unbiased way; and
- Complete documentation of all sampling steps is maintained.

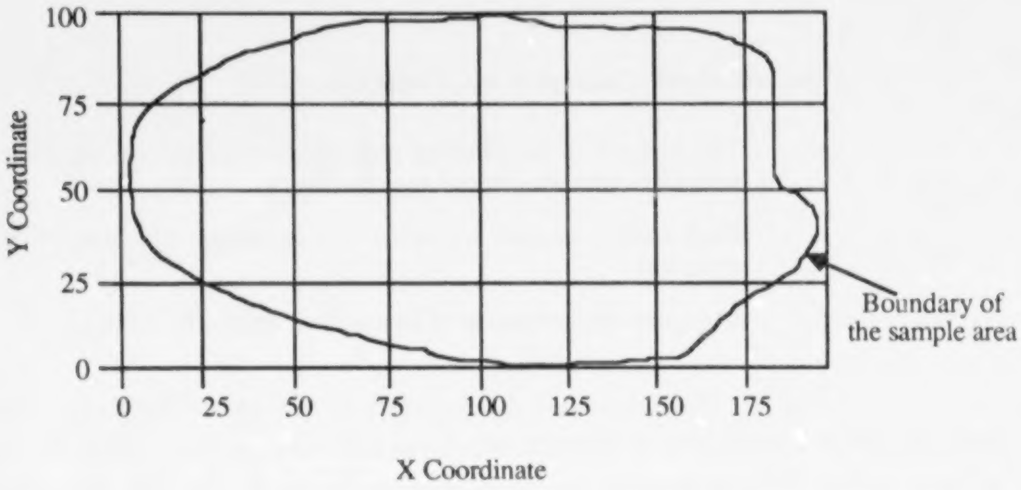
The procedures discussed in this chapter assume that the sampling plan has been selected; the boundaries of the waste site, the sample areas, and any strata have been defined; a detailed map of the waste site is available; and the sample size is known. Sample size determination is discussed in Chapters 6, 7, 8, and 9. Also, if sequential sampling or hot spot searches are planned, the reader should refer to Chapters 8 and 9, respectively, for additional guidance on field sampling.

5.1 Determining the General Sampling Location

Locating the soil samples is accomplished using a detailed map of the waste site with a coordinate system to identify sampling locations. Recording and automation of station-specific data should retain coordinate information, especially if geostatistical manipulations are performed (see Chapter 10) or a geographic information system will be used.

Soil sample locations will be identified by X and Y coordinates within the grid system. It is not necessary to draw a grid for the entire waste site; it is only necessary to identify the actual coordinates selected. Figure 5.1 is an example of a map with a coordinate system. In this example, the origin of the coordinate system is at the lower lefthand corner of the map; however, this may not be true for coordinate systems based on measurements from a reference point on the ground, i.e., a benchmark or a standard coordinate system such as latitude and longitude.

Figure 5.1 Map of a Sample Area with a Coordinate System



The boundaries of the sample areas (areas within the site for which separate attainment decisions are to be made) and strata within the sample areas (if stratified sampling is required) should be shown on the map. The map should also include other important features that will be useful in identifying sample locations in the field.

Accurate location of sampling points can be expensive and time consuming. Therefore, a method is suggested which uses the coordinate system to identify the general area within which the soil sample is to be collected, followed by a second stage of sampling, described in section 5.5, to identify the sample point accurately.

The X and Y coordinates of each sample location must be specified. This distance between coordinates on each axis represents a reasonable accuracy for measuring distance in the field, and is represented by M. If distances can be measured easily to within 2 m, but not to within 1 meter, the coordinates should be provided to the nearest 2 m (M = 2 m). The sampling coordinates can be identified with greater accuracy when the distances to be measured between reference points are short, the measuring equipment is accurate or easy to use, or there are few obstructions to line-of-sight measuring such as hills, trees, or bushy vegetation. For example, the location within a small lagoon, say, 30 by 30 m, can be established to within 5 cm. On the other hand, in a 10 hectare field it may only be reasonable to identify a location to within 10 m.

5.2 Selecting the Sample Coordinates for a Simple Random Sample

A random sample of soil units within the sample area or stratum will be selected by generating a series of random (X,Y) coordinates, finding the location in the field associated with these (X,Y) coordinates, and following the field procedures described in section 5.5 for collecting soil samples. If the waste site contains multiple sample areas and/or strata, the same procedure described above is used to generate random pairs of coordinates with the appropriate range until the specified sample size for the particular portion of the site has been met. In other words, a separate simple random sample of locations should be drawn for each sample area or stratum. To simplify the discussion, the procedures below discuss selection of a random sample in a sample area.

The number of soil samples to be collected must be specified for each sample area. In what follows, the term n_f will be used to denote the number of samples to be collected in the sample area.

To generate the n_f random coordinates (X_i,Y_i), $i = 1$ to n_f, for the sample area, determine the range of X and Y coordinates that will completely cover the sample area. These coordinate ranges will define a rectangle that circumscribes the sample area. Let the coordinate ranges be X_{min} to X_{max} and Y_{min} to Y_{max}. Thus, the point (X_{min}, Y_{min}) represents the lower lefthand corner of the rectangle, and (X_{max}, Y_{max}) represents the upper righthand corner of the rectangle. The n_f sample coordinates (X_i,Y_i) can be generated using a random number generator and the steps described in Box 5.1. Box 5.2 gives an example of generating random sample locations.

Box 5.1
Steps for Generating Random Coordinates That Define Sampling Locations

1) Generate a set of coordinates (X,Y) using the following equations:

$$X = X_{min} + (X_{max} - X_{min})*RND \qquad (5.1)$$

$$Y = Y_{min} + (Y_{max} - Y_{min})*RND \qquad (5.2)$$

RND is the next unused random number between 0 and 1 in a sequence of random numbers. Random numbers can be obtained from calculators, computer software, or tables of random numbers.

2) If (X,Y) is outside the sample area, return to step 1 to generate another random coordinate; otherwise go to step 3.

3) Define (X_i,Y_i) using the following steps:

Round X to the nearest unit that can be located easily in the field (see section 5.1); set this equal to X_i

Round Y to the nearest unit that can be located easily in the field (see section 5.1); set this equal to Y_i.

4) Continue to generate the next random coordinate, (X_{i+1},Y_{i+1}).

Box 5.2
An Example of Generating Random Sampling Locations

To illustrate the selection of simple random sample of locations, assume that seven soil units will be selected from the site in Figure 5.2. Pairs of random numbers (one X coordinate and Y coordinate for each pair) identify each sample point. X will be measured on the map's coordinate system in the horizontal direction and Y in the vertical direction. It is assumed for this example that selected coordinates can be identified to the nearest meter. The first number of pair, X_i, must be between 0 and 190 (i.e., $X_{min} = 0$ and $X_{max} = 190$) and the second, Y_i, between 0 and 100 ($Y_{min} = 0$ and $Y_{max} = 100$) for this example. If the X and Y coordinates for any pair identify a location outside the area of interest, they are ignored and the process is continued until the sample size n_f has been achieved.

X Y pair	Random X coordinate	Random Y coordinate
1	67	80
2	97	4
3	190	88 (outside of sample area)
4	17	15 (outside of sample area)
5	94	76
6	123	49
7	25	52
8	35	39
9	152	14

It took nine attempts to secure seven coordinates that fall within the sample area. The randomly selected coordinates for pairs 3 and 4 fall outside the waste site and are to be discarded. The remaining seven locations are randomly distributed throughout the site.

These locations can now be plotted on the map, as shown in Figure 5.2.

5.3 Selecting the Sample Coordinates for a Systematic Sample

A square grid and a triangular grid are two common patterns used in systematic or grid sampling. These patterns are shown in Figure 5.3. Note that the rows of points in the triangular grid are closer (.866L) than the distance between points in a row (L) and that the points in every other row are offset by half a grid width.

Figure 5.2 Map of a Sample Area Showing Random Sampling Locations

Locations of the random samples are indicated by a •. The numbers reference the XY pairs in Box 5.2.

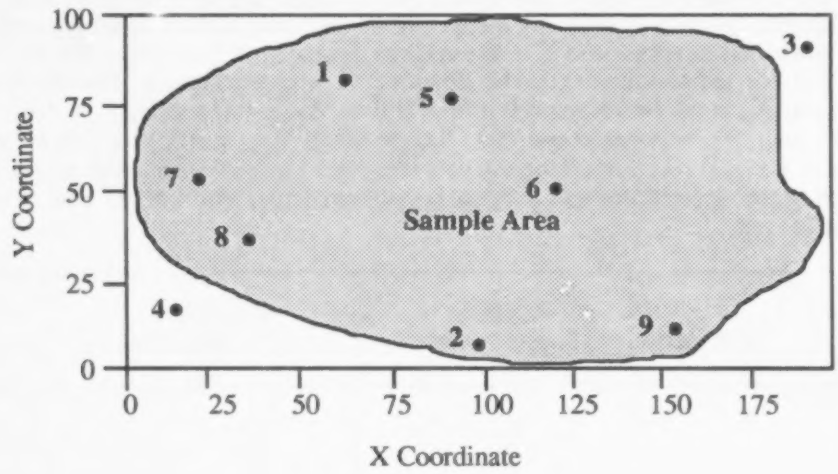
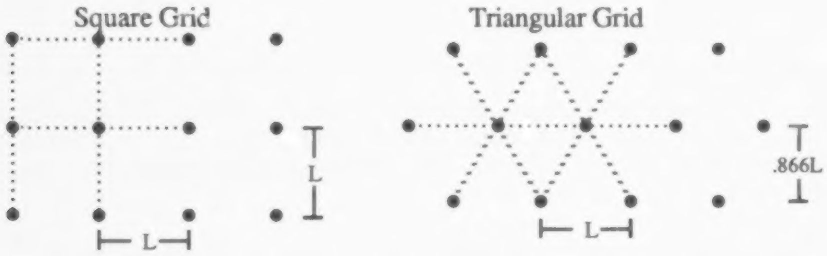


Figure 5.3 Examples of a Square and a Triangular Grid for Systematic Sampling



The size of the sample area must be determined in order to calculate the distance, L, between the sampling locations in the systematic grid. The area can be measured on a map using a planimeter. The units of the area measurement (such as square feet, hectares, square meters) should be recorded.

Denote the surface area of the sample area by A. Use the equations in Box 5.3 to calculate the spacing between adjacent sampling locations.

Box 5.3
Calculating Spacing Between Adjacent Sampling Locations

for the Square Grid in Figure 5.3

$$L = \sqrt{\frac{A}{n_f}} \qquad (5.3)$$

for the Triangular Grid in Figure 5.3

$$L = \sqrt{\frac{A}{.866 n_f}} \qquad (5.4)$$

The distance between adjacent points, L, should be rounded to the nearest unit that can be easily measured in the field.

After computing L, the actual location of one point in the grid should be chosen by a random procedure. First, select a random coordinate (X,Y) following the procedure in Box 5.1. Using this location as one intersection of two gridlines, construct gridlines running parallel to the coordinate axes and separated by a distance L. The sampling locations are the points at the intersections of the gridlines that are within the sample area boundaries. Figure 5.4 illustrates this procedure. Using this procedure, the grid will always be oriented parallel to the coordinate axes. The grid intersections that lie outside the sample area are ignored. There will be some variation in sample size, depending on the location of the initial randomly drawn point. However, the relative variation in number of sample points becomes small as the number of desired sample points increases. For unusually shaped sample areas (or strata), the number of sample points can vary considerably from the desired number.

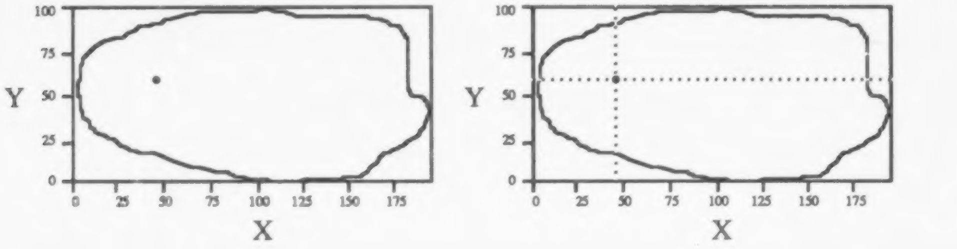
The coordinates for the sample points will be all coordinates (X_i, Y_i) such that:

- (X_i, Y_i) is inside the sample area or stratum;
- $X_i = X + j*L$, for some positive or negative integer j, and;
- $Y_i = Y + k*L$, for some positive or negative integer k.

Figure 5.4 Locating a Square Grid Systematic Sample

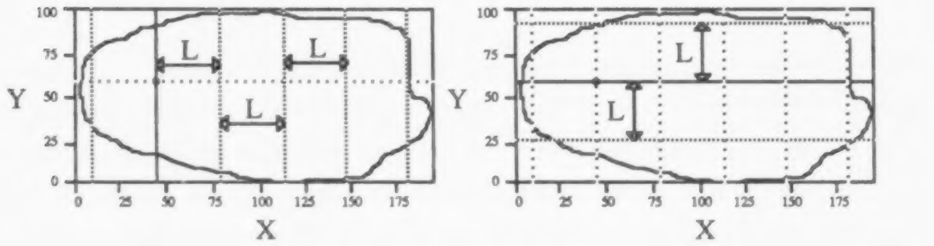
(1) Select initial random point.

(2) Construct coordinate axis going through initial point.


100
75
Y 50
25
0
0 25 50 75 100 125 150 175
X
100
75
Y 50
25
0
0 25 50 75 100 125 150 175
X

(3) Construct lines parallel to vertical axis, separated by a distance of L.

(4) Construct lines parallel to horizontal axis, separated by a distance of L.


100
75
Y 50
25
0
L
L
L
0 25 50 75 100 125 150 175
X
100
75
Y 50
25
0
L
L
0 25 50 75 100 125 150 175
X

Box 5.4 and Figure 5.5 give an example of locating systematic coordinates and the resultant sampling locations plotted on a map of the site.

Box 5.4
Locating Systematic Coordinates

Using the map in Figure 5.1 and a planimeter, the area of the sample area is determined to be 14,025 sq. m. If the sample size is 12, the spacing between adjacent points is:

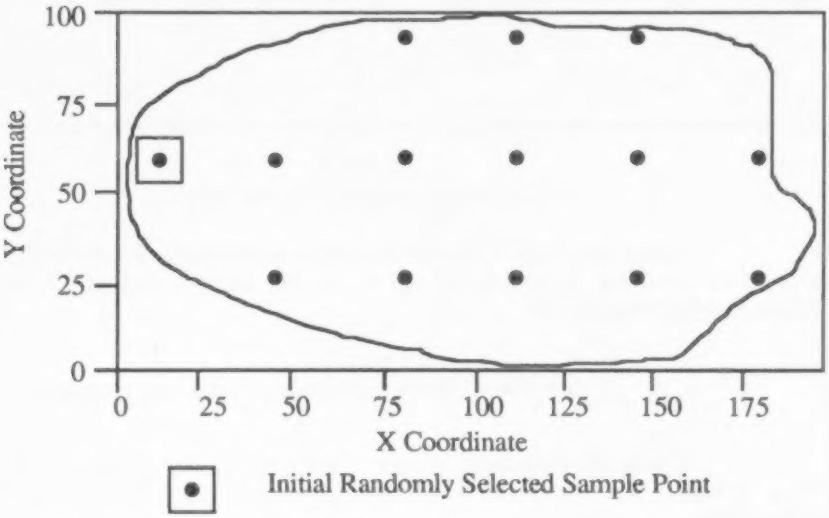
$$L = \sqrt{\frac{A}{n_f}} = \sqrt{\frac{14025}{12}} = 34 \text{ m, rounded to the nearest meter}$$

Using the procedure in Box 5.1, a random coordinate (X,Y) = (11,60) is generated. Starting from this point, the following sampling points can be calculated:

		(79,94)	(113,94)	(147,94)	
(11,60)	(45,60)	(79,60)	(113,60)	(147,60)	(181,60)
	(45,26)	(79,26)	(113,26)	(147,26)	(181,26)

These points are shown in Figure 5.5. The intended sample size was 12; however, because of the random selection process and the irregularity of the sample area boundary, there are 14 sample points within the sample area. A sample will be collected at all 14 locations.

Figure 5.5 Map of a Sample Site Showing Systematic Sampling Locations


100
75
50
25
0
Y Coordinate
0
25
50
75
100
125
150
175
X Coordinate
Initial Randomly Selected Sample Point

5.3.1 An Alternative Method for Locating the Random Start Position for a Systematic Sample

An alternative method may be used to locate the random start position for a systematic triangular grid sample (J. Barich, Pers. Com., 1988). This approach, as detailed in Box 5.5, determines a random start location by choosing a random angle A and a random distance Y from point X. This approach is useful under circumstances where a transit and stadia rod are available for turning angles, measuring distances, and establishing transects. This method is essentially equivalent to the method described above.

Box 5.5
Alternative Method for Locating the Random Start Position for a Systematic Sample

Figure 5.6 and the following steps explain how to implement the sequence.

1) Establish the main transect with endpoints X and X' using any convenient reference line (e.g., established boundary). Notice that the transect X-X' must be longer than the line indicated in Figure 5.6 in order to site all of the transects that intersect the sample area.

2) Randomly choose a point Y between X and X'.

3) Randomly choose an angle A between 0° and 90°.

4) Locate transect with endpoints Y and Y', A degrees from transect X and X'. If this transect intersects the boundary of the sample area, mark the transect.

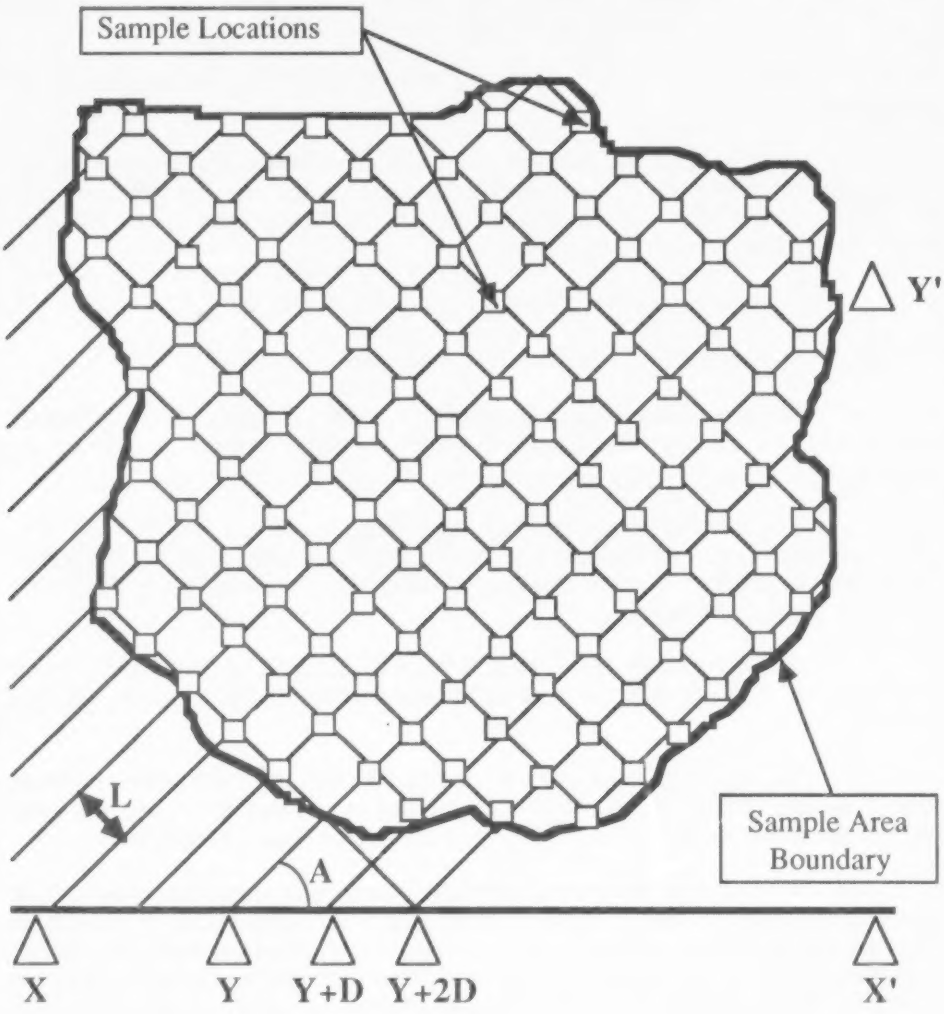
5) Locate another transect beginning at point Y and 90° +A (i.e., perpendicular) from that transect that intersects the boundary of the sample area; then mark the transect Y-Y'. If this transect intersects the boundary of the sample area then mark the transect.

6) Move away from point Y on transect X-X' a distance D, where D=L/sin(A). L is the desired interval between sampling points along the grid pattern.

7) At the point D units away from Y, establish two more transects: one A degrees from transect X-X' and parallel to transect Y-Y', and the other 90°+A degrees from X-X' also beginning at the point D units from point Y.

8) Continue to move intervals of distance D along the transect X-X' until two transects intersect within the boundary of the sample area. Establish the first sample location at that point. Then measure along that transect from the first sampling location a distance of L and establish more transects and grid points using the approach described in the previous method for systematic samples.

Figure 5.6 Method for Positioning Systematic Sample Locations in the Field


Sample Locations
Y'
Sample Area Boundary
L
A
X
Y
Y+D
Y+2D
X'

Where **D = L/sin A**

Y is chosen randomly

A is chosen randomly

L is determined from sample size calculations

☐ is a physical sampling location

5.4 Extension to Stratified Sampling

The extension of these procedures to stratified sampling is straightforward. Each stratum is sampled separately using the methods discussed above. Different random sequences (or random numbers for locating the grids) should be used in each stratum within the sample area. The sampling approach chosen for one stratum does not have to be used in another stratum. For example, if a sample area is made up of a small waste pile and a large 200-acre hillside, then it would be possible to use systematic sampling for the hillside and random sampling for the waste pile.

5.5 Field Procedures for Determining the Exact Sampling Location

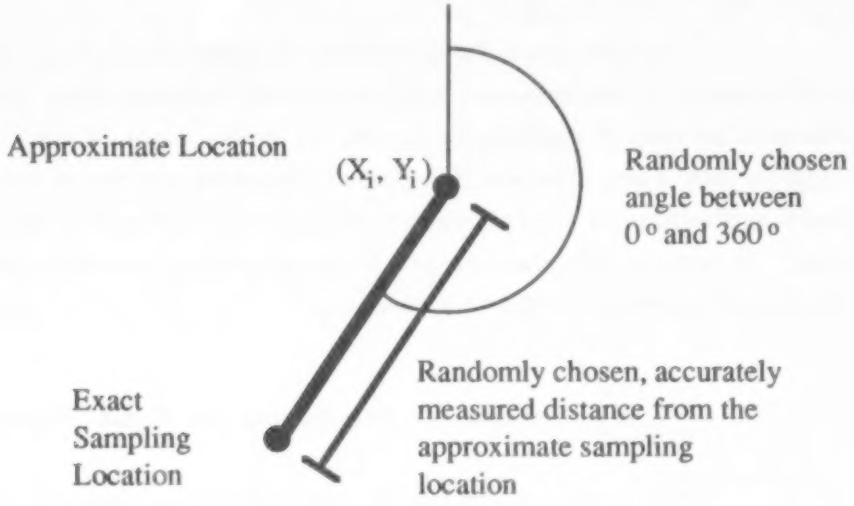
The grid points specified for the coordinate system or other reference points (e.g., trees, boulders, or other landmarks) provide the starting point for locating the sample points in the field. The location of a sample point in the field will be approximate because the sampling coordinates were rounded to distances that are easy to measure, the measurement has some inaccuracies, and there is judgment on the part of the field staff in locating the sample point.

A procedure to locate the exact sample collection point is recommended to avoid subjective factors that may affect the results. Without this precaution, subtle factors such as the difficulty in collecting a sample, the presence of vegetation, or the color of the soil may affect where the sample is taken, and thus bias the results.

To locate the exact sample collection point in the field, use one of the following procedures (or a similar procedure) to move from the location identified when measuring from the reference points to the final sample collection point. In the methods below, M is the accuracy to which distances can be easily measured in the field.

- Choose a random compass direction (0 to 360 degrees or N, NE, E, SE, etc.) and a random distance (from zero to M meters) to go to the sample location (as illustrated in Figure 5.7).
- Choose a random distance (from -M meters to M meters) to go in the X direction and a random distance (from -M meters to M meters) to go in the Y direction, based on the coordinate system.

Figure 5.7 An Example Illustration of How to Choose an Exact Field Sampling Location from an Approximate Location


Approximate Location
(X_i, Y_i)
Randomly chosen angle between 0° and 360°
Exact Sampling Location
Randomly chosen, accurately measured distance from the approximate sampling location

For either of these procedures, the random numbers can be generated in the field using a hand-held calculator or by generating the random numbers prior to sampling. The sample should be collected as close to this exact sampling location as possible.

5.6 Subsampling and Sampling Across Depth

Methods for deciding how and where to subsample a soil core are important to understand and include in a sampling plan. These methods should be executed consistently throughout the site. The field methods that are used will depend on many things including the soil sampling device, the quantity of material needed for analysis, the contaminants that are present, and the consistency of the solid or soils media that is being sampled. The details of how these considerations influence field procedures are not the subject of this discussion, but they are important and related to the discussion. More detail can be obtained in the Soil Sampling Quality Assurance User's Guide (USEPA, 1984).

This discussion describes methods for soil acquisition across depth once an exact auguring or coring position has been determined and describes how these approaches influence the interpretation of sampling results. There are several approaches that might be considered each with advantages and disadvantages; these are outlined in Figure 5.8.

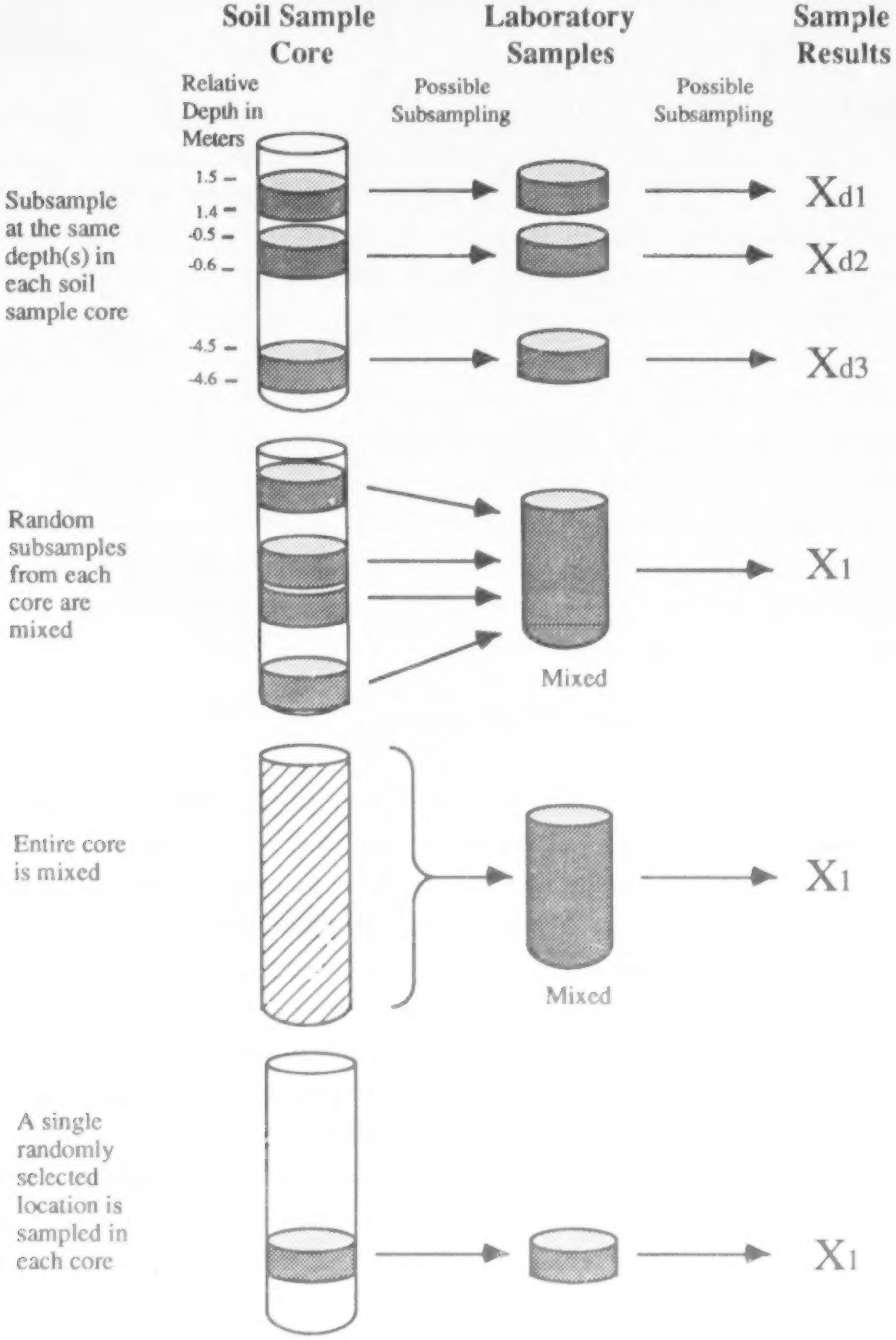
5.6.1 Depth Discrete Sampling

The first approach is to decide before sampling on an exact position or positions across depth that will be retained for analysis. For example, it may be decided that throughout the site a split spoon will be driven so that the soil within the following intervals is retained and sent to the laboratory for separate analysis: at elevations 1.5 m to 1.4 m, -0.5m to -0.6 m, and -4.5 m to -4.6 m (relative to a geodetic or site standard elevation). The size of the interval would depend on the volume required by the laboratory. In this example, all the soils material within each interval is extracted and analyzed. Advantages of this approach are that each depth can be considered a different sample area and conclusions regarding the attainment of cleanup standards can be made independently for each soil horizon. This is also a preferred method when the presence of volatiles in the soils media prevents the application of compositing methods.

5.6.2 Compositing Across Depth

Other approaches to sample acquisition within a core are based on compositing methods. Compositing methods are generally to be approached with caution unless the statistical parameter of interest is the mean concentration. If the mean is the statistic of interest, then the variance of the mean contributed by differences in location across the site from composited samples will be lower than the same variance associated with the mean from noncomposited samples. However, compositing will restrict the evaluation of the proportion of soil above an established cleanup standard because of the physical averaging that occurs in the compositing process. Clearly compositing is not recommended if the compositing process will influence the mass of material in the sample as in the case of volatile organics within a soils matrix. Numerous authors have contributed to the understanding of the effects of compositing (Duncan, 1962; Elder et al., 1980; Rohde, 1976; Schaeffer and Janardan, 1978; and Schaeffer et al., 1980), and these references or a statistician should be consulted if complicated compositing strategies are planned.

Figure 5.8 Subsampling and Sampling Across Depth


Soil Sample Core
Laboratory Samples
Sample Results
Relative Depth in Meters
Possible Subsampling
Possible Subsampling
1.5
1.4
-0.5
-0.6
-4.5
-4.6
Subsample at the same depth(s) in each soil sample core
X_{d1}
X_{d2}
X_{d3}
Random subsamples from each core are mixed
Mixed
X_1
Entire core is mixed
Mixed
X_1
A single randomly selected location is sampled in each core
X_1

Under one compositing method, segments of the soil core are retained from randomly or systematically identified locations. Then only the sampled portions are homogenized and then subsampled. Another approach calls for retaining the entire core and homogenizing all of the material and then subsampling. The latter approach is preferred from a statistical point of view because the subsampling variance will be lower. However, the second method may present difficulties if the soil samples are obtained to considerable depth or by split spoon. In these situations, it is clearly not reasonable or cost effective to acquire a core from the entire soil profile. On the other hand, if a hand-held core or continuous coring device such as a vibra-corer is being used, then homogenization of the entire core may be possible. In general, large amounts of material, material that is difficult to manipulate because of its physical properties, material containing analytes that will volatilize, or hazardous soil make thorough mixing more difficult, which may eventually defeat the positive features associated with homogenization of the entire core.

5.6.3 Random Sampling Across Depth

A final approach involves randomly sampling a single location within each core. At first, this approach appears to have many difficulties, but if the interest is in verifying that the proportion of soil above a cleanup standard is low, this approach will work quite well.

Suppose that an in situ soils stabilization method was used to treat all of the overburden soils within a former lagoon. The treatment was previously found to yield effective and homogeneous results over depth and space. It would clearly not be appropriate to sample at a single depth of, say, 3m. Since depth homogeneity is expected, it may also not be necessary to evaluate several specific depths by sampling 1-m, 3-m, 7-m, and 15-m horizons in each boring. Finally and most importantly, it would not be recommended to perform compositing because the statistical parameter of interest is the proportion of soil at the site above the cleanup standard and not the mean concentration.

In this situation it may be useful to pick a random depth at each location. In this way, many depths will be represented across the lagoon. Also, cost may be reduced because at many locations the auger will not have to drill to bedrock because the sample will be obtained from a random location that, in some samples, will be near the surface.

5.7 Quality Assurance/Quality Control (QA/QC) in Handling the Sample During and After Collection

Data resulting from a sampling program can only be evaluated and interpreted with confidence when adequate quality assurance methods and procedures have been incorporated into the design. An adequate quality assurance program requires awareness of the sources of error associated with each step of the sampling effort.

A full discussion of this topic is beyond the scope of the document; however, the implementation of a QA program is important. For additional details, see *Soil Sampling Quality Assurance User's Guide* (USEPA, 1984), Brown and Black (1983), and Garner (1985).

5.8 Summary

Locating soil samples is accomplished using a detailed map of the waste site with a coordinate system to identify sampling locations. The boundaries of the sample areas (areas within the site for which separate cleanup verification decisions are to be made) and strata within the sample areas should be shown on the map. It is not necessary to draw a grid for the entire waste site, only to identify the actual coordinates selected.

A random sample of soil units within the sample area or stratum will be selected by generating a series of random (X,Y) coordinates and identifying the location associated with these coordinates.

When selecting the sample coordinates for a systematic sample, two common patterns of systematic or grid samples are a square grid and a triangular grid. Various methods can be used to select a systematic sample; however, the most important point is that one of the systematic sample locations must be identified randomly.

A separate random or systematic sample is selected for each sample area. In addition, the extension of these procedures to stratified sampling is straightforward. Each stratum is sampled separately. The sampling approach chosen for one stratum, or sample area does not have to be used in another stratum.

Once a horizontal position is chosen, the method of acquiring samples across depth must be decided. Methods for subsampling and sampling across depth should be executed consistently throughout the site. The methods discussed are:

- Depth discrete sampling;
- Compositing across depth; and
- Random sampling across depth.

6. DETERMINING WHETHER THE MEAN CONCENTRATION OF THE SITE IS LESS THAN A CLEANUP STANDARD

This chapter describes statistical procedures for determining whether the mean concentration in the sample area attains the cleanup standard. Testing whether the mean attains the cleanup standard is appropriate if the mean (or average) concentration is of particular interest and if the higher concentrations found in limited areas are not of concern. If the median concentration or the more extreme concentrations (e.g., the concentration for which 95 percent of the site is lower and 5 percent of the site is higher) are of interest, then see Chapter 7 for appropriate statistical techniques.

The statistical procedures given in this chapter for deciding if the mean concentration attains the cleanup standard are called "parametric" procedures. They usually require certain assumptions about the underlying distribution of the data. Fortunately, the procedures perform well even when these assumptions are not strictly true, and thus they are applicable in many different field conditions (see Conover, 1980).

The following topics--determination of sample size; calculation of the mean, standard deviation, and confidence interval; and deciding if the sample area attains the cleanup standard--are discussed for each of the following sample plans in the sections indicated:

- Simple random sampling (section 6.3);
- Stratified random sampling (section 6.4); and
- Systematic sampling (section 6.5).

6.1 Notation Used in This Chapter

The following notation is used throughout this chapter:

Cs The cleanup standard relevant to the sample area and the contaminant being tested.

μ	The “true” but unknown mean contaminant concentrations across the sample area, the population mean.
H_0	The null hypothesis, which is assumed to be true in the absence of significant contradictory data. When testing the mean, the null hypothesis is that the sample area does not attain the cleanup standard: $H_0: \mu \geq Cs$.
α	The desired false positive rate for the statistical test. The false positive rate for the statistical procedure is the probability that the sample area will be declared to be clean when it is actually dirty.
H_1	The alternative hypothesis, which is declared to be true only if the null hypothesis is shown to be false based on significant contradictory data. When testing the mean, the alternative hypothesis is that the sample area attains the cleanup standard: $H_1: \mu < Cs$.
μ_1	The value of μ under the alternative hypothesis for which a specified false negative rate is to be controlled ($\mu_1 < \mu$).
β	The false negative rate for the statistical procedure is the probability that the sample area will be declared to be dirty when it is actually clean and the true mean is μ_1. The desired sample size n_d is selected so that the statistical procedure has a false negative rate of β at μ_1.
n_d	The desired sample size for the statistical calculations.
n	The final sample size, i.e., the number of data values available for statistical analysis including the concentrations that are below the detection level.
x_i	The contaminant concentration measured for soil sample i, i = 1 to n. For measurements reported as below detection, x_i = the detection limit. See section 2.5.2 for more details.

6.2 Calculating the Mean, Variance, and Standard Deviation

For many purposes in this chapter it is necessary to calculate the mean, variance, or standard deviation for a sample of data. The basic formulas are provided in Box 6.1 for use in later sections.

Box 6.1
Calculating Sample Mean, Variance, Standard Deviation, and Coefficient of Variation

If the data are a random sample of n observations (i.e., the sample size is n), designate the data as $x_1, x_2 \ldots, x_i, \ldots$ to x_n. The sample mean (or average), indicated by $\bar{x}$, is calculated as:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (6.1)$$

The formula for the sample variance, s^2, is:

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} \qquad (6.2)$$

The formula for the standard deviation is:

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}} \qquad (6.3)$$

The formula for the coefficient of variation is:

$$cv = \frac{s}{\bar{x}} \qquad (6.4)$$

The standard deviation provides a measure of the variability of the sample data. In particular, it is used to obtain estimates of standard errors and confidence limits.

Degrees of freedom, denoted by df, provide a measure of how much information the variance or standard deviation is based on. The variance and the standard deviation calculated above for simple random samples have "n-1 degrees of freedom." The degrees of freedom are used in calculating confidence intervals and performing hypothesis tests to determine whether the sample area has attained the cleanup standard.

6.3 Methods for Random Samples

Methods in this section are applicable when the criterion for deciding whether the site attains the cleanup standard is based on the mean concentration and the samples are collected using simple random sampling. The steps involved in the data collection and analysis are:

- Determine the required sample size (section 6.3.2);
- Identify the locations within the site from which the soil samples are to be collected and collect the physical samples for analysis (Chapter 5);
- Perform appropriate statistical analysis using the procedures described in section 6.3.3 and on the basis of the decision rule given in section 6.3.4, decide whether the site requires additional cleanup.

6.3.1 Estimating the Variability of the Chemical Concentration Measurements

Before sample collection, determine the number of samples needed to achieve the desired confidence in the findings. The number of soil samples depends on the anticipated variability of the soil measurements. Therefore, an estimate of the standard deviation of the underlying contamination levels must be obtained. The true value of the standard deviation is denoted by the Greek letter **sigma,** σ. Estimation of σ is discussed in the next section.

To estimate the required sample size, some information about the standard deviation, σ (or equivalently the variance σ^2), is needed. Unfortunately, the standard deviation is usually unknown, and steps must be taken to estimate this quantity for the purpose of determining sample size. The symbol "^" is used to denote that $\hat{\sigma}$ is an estimate of σ. In practice, $\hat{\sigma}$ is either obtained from prior data or by conducting a small preliminary investigation such as a pilot-scale treatability study. Cochran (1977) discusses aspects of determining a preliminary value for $\hat{\sigma}$.

6.3.1.1 Use of Data from a Prior Study to Estimate σ

If there are data on contamination levels for the site under investigation from a previously selected sample of soil units or a treatability study, this information can be used to obtain $\hat{\sigma}$. Note that the characteristics of physical samples used in the previous study should be roughly the same as those planned for the present evaluation. For best results, the sample from the prior study should be a simple random sample. If not, the sample should at least be "representative" in the sense that the measurements are distributed evenly across the cleanup area. In particular, measurements that tend to be located within a specific subarea would generally be inappropriate for estimating the variability across the entire area.

To obtain $\hat{\sigma}$ from the existing sample, calculate the variance of the chemical observations. It is best to have at least 20 observations for the variance calculations. The sample standard deviation, s, can be calculated using equation (6.3) in Box 6.1. Use the calculated value of s for $\hat{\sigma}$.

6.3.1.2 Obtain Data to Estimate σ After a Remedial Action Pilot

This approach will be best implemented as part of a pilot scale treatability study.

1) Using the sampling procedures described in Chapter 5, select a preliminary (simple random) sample of $n_1 = 20$ soil units. Determine the concentrations for these 20 units.

2) From this preliminary sample, compute the standard deviation, s, of the contaminant levels. Using s for $\hat{\sigma}$, determine the required sample size, n, using equation (6.6) in Box 6.3.

3) If the sample size determined is less than or equal to 20, proceed with the statistical analysis as outlined in sections 6.3.2, 6.3.3, and 6.3.4, using the preliminary sample as the complete sample. Otherwise, select enough additional soil samples so that the preliminary sample plus the additional samples add up to the required sample size. In this case, the results for the initial sample and the supplement should be combined for the statistical analysis.

6.3.1.3 An Alternative Approximation for $\hat{\sigma}$

If there are no existing data to estimate σ, and a preliminary study is not feasible, a crude approximation for $\hat{\sigma}$ can be obtained. The approximation is based on speculations and judgments concerning the range within which the soil measurements are likely to fall. The approximation is based on virtually no data, so the sample sizes computed from these approximations may not satisfy the specified level of precision. Consequently, it should only be used if no other alternative is available.

The approximation described in Box 6.2 uses the range of possible soil measurements (i.e., the largest possible value minus the smallest value). The range provides a measure of the variability of the data. Moreover, if the frequency distribution of the soil measurements of interest is approximately bell-shaped, then over 99 percent of the measurements can be expected to lie within three standard deviations of the mean.

Box 6.2
An Alternative Approximation for $\hat{\sigma}$

An estimate of σ is given by:

$$\hat{\sigma} = \text{RANGE}/6 \qquad (6.5)$$

Where RANGE = the expected spread between the smallest and largest values.

6.3.2 Formulae for Determining Sample Size

The equations for determining **sample size** require the specification of equations 6.6 and 6.7, given in Box 6.3 and the following quantities: cleanup standard (Cs), the mean concentration where the site should be declared clean with a high probability (μ_1), the false positive rate (α), the false negative rate (β), and the standard deviation ($\hat{\sigma}$).

Box 6.3
Formulae for Calculating the Sample Size
Needed to Estimate the Mean

$$n_d = \hat{\sigma}^2 \left\{ \frac{z_{1-\beta} + z_{1-\alpha}}{Cs - \mu_1} \right\}^2 \qquad (6.6)$$

where $z_{1-\beta}$ and $z_{1-\alpha}$ are the critical values for the normal distribution with probabilities of $1-\alpha$ and $1-\beta$ (Table A.2).

The sample size may also be written in the following equivalent form:

$$n_d = \frac{(z_{1-\beta} + z_{1-\alpha})^2}{\tau^2} \quad \text{where } \tau = \frac{(Cs - \mu_1)}{\hat{\sigma}}. \qquad (6.7)$$

The term τ (Greek letter tau) expresses the difference in units of standard deviation. For convenience, the values of n as computed from this formula are given in Table A.6 for selected values of α, β, and τ.

Box 6.4 gives an example of calculating sample size.

Box 6.4
Example of Sample Size Calculations

Suppose it is desirable to verify cleanup when the mean concentration is .1 ppm below the cleanup standard of .5 ppm ($Cs = .5$, $\mu_1 = .3$) with a power of .80 (i.e., $\beta = .20$). Also suppose $\sigma = .43$, $\alpha = .05$, and 99 percent of the sample points will result in analyzable samples, then

$$\tau = \frac{(Cs - \mu_1)}{\sigma} = \frac{(.5 - .3)}{.43} = .465$$

From Table A.6 with $\beta = .20$, $\alpha = .05$, and $\tau = .465$, the desired sample size is between 25 and 30. Using linear interpolation gives a sample size of about 29. From Table A.2,

$$z_{1-\alpha} = 1.645,\ z_{1-\beta} = 0.842.$$

Using formula 6.7,

$$n_d = \frac{(z_{1-\beta} + z_{1-\alpha})^2}{\tau^2} = \frac{(.842 + 1.645)^2}{.465^2} = 28.6$$

and

$$n_f = \frac{n_d}{R} = \frac{28.6}{.99} = 28.9.$$

Rounding up, the sample size is 29.

Box 6.5 gives an example of determining sample size for testing the mean using power curves.

Box 6.5
Example: Determining Sample Size for Testing the Mean Using the Power Curves

At a former wood processing plant it is desirable to determine if the average concentrations of PAH compounds in the surface soil are below 50 ppm (the cleanup standard Cs). The project managers have decided that the dangers from long-term exposure can be reasonably controlled if the mean concentration in the sample area is less than the cleanup standard. The false positive rate for the test is to be at most 5 percent (i.e., $\alpha = .05$). The coefficient of variation of the data is thought to be about 1.2. After reviewing the power curves in Figure A.2 and the approximate sample sizes for random sampling, the managers decide:

1) While it would be desirable to have a test with power curves similar to curves E and F, the samples sizes of more than 100 will cost too much.

2) Power curves A, B, and C have unacceptably low power when the mean concentration is roughly 75 percent of the cleanup standard (i.e., 37 ppm), the expected mean based on a few preliminary samples.

3) Thus the test should have power similar to that in curve D.

Based on the specifications above and the table at bottom of the Figure A.2, the information needed to calculate the sample size is:

$$\alpha = .05;$$

$$\beta = .20; \text{ and}$$

$$\mu_1 = Cs * .69 = 34.5 \text{ ppm.}$$

These values can be used to determine the sample size using the equations described earlier.

If the sample size has been specified in advance, perhaps based on cost considerations, Figures A.1 through A.4 can be used to determine the approximate shape of the power curve for the associated test. See Box 6.6 for an example.

Box 6.6
Determining the Approximate Power Curve
for a Specified Sample Size

Suppose that after review of the budget and analytical costs, the managers had decided that 40 samples would be collected. What is the approximate shape of the power curve for the associated test assuming $\alpha = .05$, $\beta = .20$, and a systematic sample is used?

Based on previous samples the managers believe that the coefficient of variation of the concentration measurement will be around 1.1. Assuming that a systematic sample will behave statistically like a random sample (a reasonable assumption of a site which has been cleaned up) and looking at the bottom of Figure A.2 at the sample sized for testing the mean:

1) If the cv were 1.0, the power curve for a sample size of 40 would be between curves C (sample size = 34) and curve D (sample size = 65), and closer to curve C.

2) If the cv were 1.5, the power curve for sample size of 40 would be between curves A (sample size = 25) and curve B (sample size = 43), and closer to curve B.

3) Since the cv is about 1.2, the power curve for the test will be between curves B and C.

6.3.3 Calculating the Mean, Standard Deviation, and Confidence Intervals

This section describes the computational procedures used to calculate the mean concentration and related quantities necessary to evaluate attainment of the cleanup standard based on a random sample. For concentrations below the detection limit, as discussed in section 2.5.2, substitute the detection limit in the calculations below.

The mean of the sampling data is an estimate of the mean contamination of the entire sample area, but does not convey information regarding the reliability of the estimate. Through the use of a "confidence interval," it is possible to provide a range of values within which the true mean is located.

The formula for an **upper *one-sided* 100(1-α) percent confidence limit** around the population mean is presented in Box 6.7. The one-sided confidence interval should be used to test whether the site has attained the cleanup standard. The corresponding decision rule is given in section 6.3.4.

Box 6.7
Computing the Upper One-sided Confidence Limit

$$\mu_{U\alpha} = \bar{x} + t_{1-\alpha,df}\frac{s}{\sqrt{n}} \qquad (6.8)$$

where $\bar{x}$ is the computed mean level of contamination, and s is the corresponding standard deviation. The appropriate value of $t_{1-\alpha,df}$ can be obtained from Table A.1.

6.3.4 Inference: Deciding Whether the Site Meets Cleanup Standards

To determine whether the site meets a specified cleanup standard, use the upper one-sided confidence limit $\mu_{U\alpha}$, defined above in equation (6.8). Use the following rule to decide whether or not the site attains the cleanup standard:

If $\mu_{U\alpha} < Cs$, conclude that the area is clean (i.e., $\mu < Cs$).

If $\mu_{U\alpha} \geq Cs$, conclude that the area is dirty (i.e., $\mu \geq Cs$).

Box 6.8 presents an example of an evaluation of cleanup standard attainment.

Box 6.8
An Example Evaluation of Cleanup Standard Attainment

From Box 6.4 the required sample size is 29. Assume for this example that all 29 field samples were collected and analyzed. Six values were below the detection level; these values were included in the analysis at the detection limit. Based on these data, the mean is .29 with a standard deviation of .41. (Note that this gives a coefficient of variation of $\frac{.41}{.29} = 1.48$).

The upper one-sided 99 percent confidence interval goes to

$$\mu_{U\alpha} = \bar{x} + t_{1-\alpha,df}\frac{s}{\sqrt{n}} = .29 + 2.467\frac{.41}{\sqrt{29}} = .478 \text{ ppm}$$

Since $0.478 < 0.5$, there is a 99 percent confidence that the mean concentration of the sample area attains the cleanup standard of 0.5 ppm.

6.4 Methods for Stratified Random Samples

The following sections discuss methods of obtaining an overall estimate of the mean contamination from a stratified sample. The steps in data collection and analysis are:

- Determine the required sample sizes for each stratum (Chapter 6.4.1);
- Within each stratum, identify the sampling locations (Chapter 5). Collect the physical samples and send the soil samples to the laboratory for analysis;
- Perform statistical analysis using the procedures described in section 6.4.2, and, on the basis of the decision rule given in section 6.4.3, decide whether the site attains the cleanup standard.

The calculations for stratified samples require knowledge of the proportion of the surface area or volume of soil represented in each stratum. The **proportion of the volume of soil** can be calculated using the formula in Box 6.9.

Box 6.9
Calculating the Proportion of the Volume of Soil

Suppose there are L strata designated by h = 1, 2, 3, ..., L. Compute the volume of soil in stratum h as:

$$V_h = \text{Surface area of stratum h} * \text{Depth of sampling in stratum h}$$

Then the proportion of the volume in stratum h is:

$$W_h = \frac{V_h}{\sum_{h=1}^{L} V_h} \qquad (6.9)$$

6.4.1 Sample Size Determination

The determination of the appropriate sample size for each stratum is complicated. There are methods (Cochran, 1977 or Kish, 1965) for determining the "optimum" allocation, but these require considerable advance knowledge about the relative costs and variability within each strata. Consequently, general guidelines, rather than rigid rules, are given in this guidance document to assist in planning the sample sizes for a stratified sample. These guidelines are expected to cover most situations likely to occur in the field. For more complex situations, consult Cochran (1977) and a statistician.

The formulas for sample size use the following notation, where h indicates the stratum number:

n_{hd}	The desired sample size for the statistical calculations in stratum h.
n_h	The final sample size in stratum h, the number of data values available for statistical analysis including the concentrations that are below the detection level.
W_h	Proportion of the volume or area of soil in the sample area that is in stratum h.
$\hat{\sigma}_h$	The estimated standard deviation of measurements from stratum h. See section 6.3.1 on estimating $\hat{\sigma}$ within a strata or sample area. If only an overall estimate, $\hat{\sigma}$, is available, use this for all strata.

C_h	The relative cost of collecting, processing, and analyzing a soil sample in strata h. If all strata are assumed to have the same relative cost for an additional sample, it will be easiest to use $C_h = 1$ for all strata.
L	The number of strata.
x_{hi}	The reported concentration of the chemical for the i^{th} sample unit in stratum h.
h	The stratum number.

After strata are defined, it is necessary to decide how many soil units should be collected in each stratum. The recommendations below are based on the following factors:

- The physical size of the stratum;
- The cost of sampling and processing a soil unit selected from the stratum; and
- The underlying variability of the chemical concentration of the soil units in the stratum.

The "optimum" sample allocation will produce the most accurate estimate of the overall mean across strata for a fixed total cost. In Boxes 6.10 and 6.11, n_{hd} will denote the desired sample size to be selected from stratum h. Thus, for a total of L strata, the overall desired sample size is $n_d = n_{1d}+n_{2d}+...+n_{Ld}$. The method for determining the **desired sample size** is given in Box 6.10 and an example is presented in Box 6.11.

Box 6.10
Calculating Desired Sample Size for Each
Stratum of a Stratified Random Sample

The desired number of soil units to be selected from stratum h is:

$$n_{hd} = \left\{ \sum_{h=1}^{L} W_h \hat{\sigma}_h \sqrt{C_h} \right\} * \left\{ \frac{z_{1-\alpha} + z_{1-\beta}}{Cs - \mu_1} \right\}^2 * \frac{W_h \hat{\sigma}_h}{\sqrt{C_h}} \qquad (6.10)$$

Box 6.11

An Example Sample Size Determination for a Stratified Sample

A site consists of two strata ($L = 2$). Stratum 1 includes loose sand soil, while stratum 2 consists of dense hard clay soil mixed with large rocks, but the same kind of contamination is present in both strata. Three meters of soil have been excavated from both areas. Stratum 1 comprises 10 percent of the sample area ($W_1 = .10$, $W_2 = .90$). The sample and analysis costs are considerably different in the two strata. The cost of sampling and analysis in stratum 2 is estimated to be 10 times that in stratum 1 ($C_1 = 1$, $C_2 = 10$) because of the cost associated with extracting a soil core. The estimated standard deviation of the measurements, based on previous sampling, is $\hat{\sigma}_1 = 25$ in stratum 1, while in stratum 2, $\hat{\sigma}_2 = 13.1$. Using a cleanup standard of 40, $\alpha = .01$, $\mu_1 = 35$ and $\beta = .20$, the sample size in each strata can be calculated as follows:

$$\left\{\frac{z_{1-\alpha} + z_{1-\beta}}{Cs - \mu_1}\right\}^2 = \left\{\frac{2.326 + .842}{40 - 35}\right\}^2 = .401$$

$$\left\{\sum_{h=1}^{L} W_h S_h \sqrt{C_h}\right\} = (.10 * 25 * \sqrt{1}) + (.90 * 13.1 * \sqrt{10}) = 39.78$$

Using equation (6.10),

$$n_{1d} = 39.78 * .401) \frac{.10 * 25}{\sqrt{1}} = 39.9$$

and

$$n_{2d} = 39.78 * .401) \frac{.90 * 13.1}{\sqrt{10}} = 59.5$$

Rounding up, and assuming that all samples will be collected and analyzed, the final sample sizes are $n_{1f} = 40$ and $n_{2f} = 60$.

When multiple statistical tests are used, or multiple chemicals tested, use the field sample size in a stratum that is the largest field sample size for any statistical test or chemical. Although this procedure for multiple tests will always provide an adequate sample size, it may not be the most cost efficient.

6.4.2 Calculation of the Mean and Confidence Intervals

If the number of values below the detection limit is moderate, procedures and formulae presented in this chapter and in the following boxes based on the sample average are applicable. Section 2.5.2 discusses the adjustment for values below the detection limit. If the proportion of values in the data set that have values below the detection limit in any stratum is large, the procedures in Chapter 7 for testing proportions may be preferred.

Box 6.12
Formula for the Mean Concentration from a Stratified Sample

The overall mean concentration, $\bar{x}_{st}$, should be computed as:

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \left(\frac{\sum_{i=1}^{n_h} x_{hi}}{n_h} \right) \qquad (6.11)$$

or

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h \qquad (6.12)$$

The equations in Box 6.13 give the formula for the **standard error** of $\bar{x}_{st}$. The standard error is required for establishing confidence limits around the actual population mean and deciding if the site attains the cleanup standard.

Box 6.13
Formula for the Standard Error from a Stratified Sample

The standard error of $\bar{x}_{st}$, denoted by $s_{\bar{x}_{st}}$ is calculated as follows:

$$s_{\bar{x}_{st}} = \sqrt{\sum_{h=1}^{L} W_h^2 \frac{s_h^2}{n_h}} \qquad (6.13)$$

where

$$s_h^2 = \frac{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2}{n_h - 1} \qquad (6.14)$$

and

$$\bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h} \qquad (6.15)$$

The approximate **degrees of freedom** for the standard error can be calculated using the formula in Box 6.14. The degrees of freedom should be rounded to the closest integer.

Box 6.14
Formula for Degrees of Freedom from a Stratified Sample

$$df = \frac{\left(s_{\bar{x}_{st}}\right)^2}{\sum_{h=1}^{L} \frac{W_h^4 s_h^4}{n_h^2 (n_h - 1)}} \qquad (6.16)$$

The mean, standard error, and degrees of freedom are used to estimate an **upper one-sided confidence interval** with a confidence of 1-α (see Box 6.15).

Box 6.15
Formula for the Upper One-sided Confidence Interval from a Stratified Sample

Compute the upper one-sided confidence limit as:

$$\mu_{U\alpha} = \bar{x}_{st} + t_{1-\alpha,df}\, s_{\bar{x}_{st}} \qquad (6.17)$$

where $\bar{x}_{st}$ is the mean level of contamination from Box 6.12, and $s_{\bar{x}_{st}}$ is the corresponding standard error from Box 6.13. The appropriate value of $t_{1-\alpha,df}$ can be obtained from Table A.1.

The value $\mu_{U\alpha}$ is a 100(1-α) percent confidence interval for the population mean.

6.4.3 Inference: Deciding Whether the Site Meets Cleanup Standards

The test statistic to be used for testing the hypothesis that the site meets specified cleanup standards is the upper one-sided confidence $\mu_{U\alpha}$, defined above in equation (6.17). Use the following rules to decide whether or not the site attains the cleanup standard. An example illustrating the procedure is in Box 6.16.

If $\mu_{U\alpha} < Cs$, conclude that the area is clean (i.e., $\mu < Cs$).
If $\mu_{U\alpha} \geq Cs$, conclude that the area is dirty (i.e., $\mu \geq Cs$).

If the upper one-sided confidence interval of the sample is below the Cs, then there is 1-α certainty that the mean of the sample area is below the Cs and the site attains the Cs.

Box 6.16
An Example Illustrating the Determination of Whether the Mean from a Stratified Sample Attains a Cleanup Standard

Following with the example in Box 6.11, the sample area has two strata. Stratum 1 comprises only 10 percent of the total site in terms of surface area. The sample consists of 40 units from stratum 1, and 60 units from stratum 2. After the soil units were analyzed in the laboratory, it was learned that two of the units in stratum 1 were below the detection limit. Hence, the chemical concentration for these two cases was set to the minimum detectable level.

1) *Calculate stratum means:* Suppose that for the 40 data values from stratum 1 the average concentration of the chemical under study was computed to be $\bar{x}_1 = 23$ ppm. Similarly, for the 60 data values from stratum 2, suppose that the average concentration was determined to be $\bar{x}_2 = 35$ ppm.

2) *Calculate stratum variances:* Using equation (6.2) the stratum standard deviations are: $s_1 = 18.2$ and $s_2 = 20.5$. Note that the 38 observations in stratum 1 that were above the detection limit, plus the two observations that were set to the minimum detectable level, were used in the calculation of s_1.

3) *Calculate overall mean*: Since 10 percent of the site is contained in stratum 1, we have $W_1 = .10$, and $W_2 = .90$. Thus, from equation (6.11), the overall mean for the entire site is:

$$\bar{x}_{st} = W_1 \bar{x}_1 + W_2 \bar{x}_2 = (.1)(23) + (.9)(35) = 33.8 \text{ ppm.}$$

4) *Calculate standard error:* The standard error of the estimate computed from the equation (6.13) is:

$$s_{\bar{x}_{st}} = \sqrt{\frac{(.1 * 18.2)^2}{40} + \frac{(.9 * 20.5)^2}{60}} = \sqrt{5.76} = 2.40$$

5) *Calculate the degrees of freedom*: Using equation (6.16),

$$df = \frac{5.76^2}{\frac{(.1 * 18.2)^4}{40^2 * 39} + \frac{(.9 * 20.5)^4}{60^2 * 59}} = \frac{33.18}{.546} = 60.8, \text{ or } df = 61$$

6) *Calculate confidence limits:* Using equation (6.17) with $t_{.99,61} =$ 2.39 from Table A.1, the upper (one-sided) 99 percent confidence limit for the "true" population mean is given by

$$\mu_{U\alpha} = \bar{x}_{st} + t_{.99,61}\, s_{\bar{x}_{st}} = 33.8 + (2.39)(2.40) = 39.54.$$

7) *Inference*: The cleanup standard for the chemical under study is Cs = 40. Since the upper confidence limit $\mu_{U\alpha} = 39.54$ is less than Cs, we conclude that the mean concentration in the sample area attains the cleanup standard.

6.5 Methods for Systematic Samples

If systematic sampling is chosen, some changes in statistical methodology may be considered and are discussed in this section. One concern is that systematic sampling should be avoided when the pattern of contamination is likely to have a cyclical or periodic pattern across the sample area. Such a situation might occur if waste was placed in trenches, if contamination blew into windows, or if a remediation technology is used such as vacuum extraction, which creates a regular pattern caused by well induced zones of influence. In such a case, a systematic sampling pattern may capture only high (or low) values of the contaminant and therefore yield biased results. It is presumed that the likelihood of this pattern will be known in advance, and be used to create strata and the need to sample randomly.

6.5.1 Estimating Sample Size

Systematic sampling can result in an increase in the precision of the statistical estimates and a corresponding decrease in the required sample size (Cochran, 1977). Unfortunately, the possible advantages of systematic sampling are difficult to predict before the sample is collected. The sampling precision of an estimated mean from a systematic sample depends on the pattern of contamination at the site and how the systematic sample is constructed. However, the standard error of a mean based on a systematic sample will usually be comparable to or less than the standard error of a mean based on a random sample of the same size. Therefore, using the sample size formulas for a random sample when the sample was collected systematically usually will be as or more protective of human health and the environment.

Use the procedures in section 6.3.2 to determine the sample size required for a systematic sample. If other procedures for calculating sample size for a systematic sample are considered, a statistician should be consulted.

6.5.2 Concerns Associated with Estimating the Mean, Estimating the Variance, and Making Inference from a Systematic Sample

When a systematic sample is obtained, apply the same methods used for a random sample. As with a simple random sample, the simple average of the sample points is an unbiased estimate of the population average. Note, however, that the number of sample points in a systematic sample of an irregularly shaped area may vary from the targeted sample size. A smaller sample size will produce estimates that have less precision than larger samples, but will not introduce bias. The loss in precision tends to be negligible except for small sample sizes.

In general, an unbiased estimate of the standard error of a mean based on a systematic sample is not available. In the special case where contamination is distributed randomly over the sample area, unbiased estimates of the standard error can be constructed. This situation may be approximately the case after a careful cleanup has been done where the cleanup has effectively removed the contaminated soil from all of the high contamination areas or the soil is being mixed, fixed, or incinerated.

Several methods are commonly used to estimate the standard error of a mean from a systematic sample (Koop, 1976; Wolter, 1984; Tornqvist, 1963; Yates, 1981). These methods treat the systematic observations as:

- A random sample;
- A stratified sample; and
- A serpentine pattern of observations that employs a special variance calculation procedure.

It is suggested that the serpentine pattern be used with overlapping pairs of points as the principal method of estimating the standard errors in a systematic design. However, if the boundaries of the sample area are so irregular as to make this approach difficult, the stratification approach is recommended. The random sample estimate should seldom be used. These approaches are discussed below.

6.5.2.1 Treating a Systematic Sample as a Random Sample

The simplest method of estimating the standard error for a systematic sample is to use the variance formulas in Box 6.1 for a simple random sample (section 6.3). This method is appropriate if the contamination is distributed randomly across the sample area. If there are gradients of contamination, or if there are substantial contiguous areas that have higher (or lower) than average contamination, this method can be biased (Osborne, 1942). In this case, the actual standard error of the mean will, on average, be smaller than that computed from the simple random sample formulas. Thus, the sample estimate will appear to be less precise than it really is and there will be a tendency to take more observations than are necessary or to do more cleanup work than is necessary.

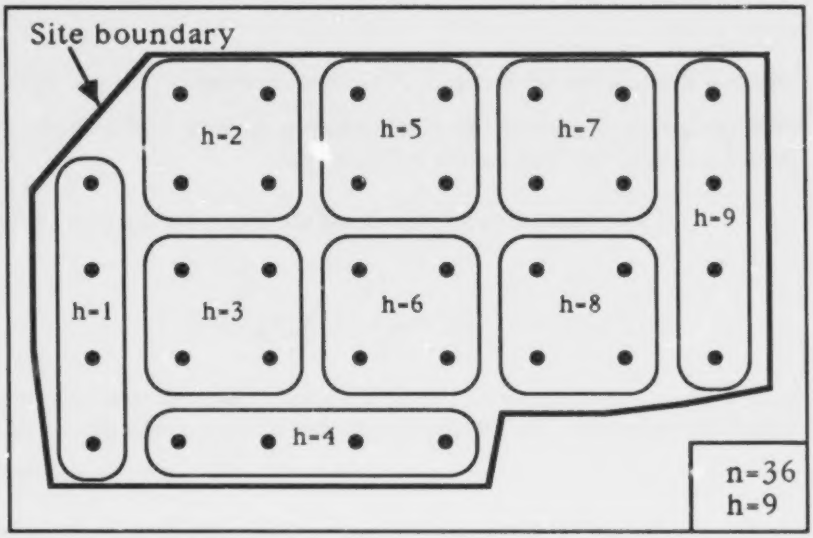
6.5.2.2 Treating the Systematic Sample as a Stratified Sample

An estimate of the standard error that is less subject to bias than the random sample estimate can be obtained by aggregating adjacent points in the systematic design into groups, and treating these groups as though they were strata (Yates, 1981) as depicted in Figure 6.1. It should be noted that this grouping can be done whether or not the sample area was previously stratified. If stratification was used, grouping for purposes of estimating standard errors would be done within strata (see Box 6.17).

A commonly used group size consists of four observations. The groups need not be the same size, but efficiency is gained if they are nearly the same size and if they are small. Points in a group should be adjacent and the groups must cover the sample area comprehensively. One must not form the groups on the basis of the observed data--this would add bias. Instead, they should be formed strictly on the basis of geographic adjacency and boundary restrictions without regard to their observed values. If the sample locations form a square grid, the recommended grouping will be four adjacent sample locations forming a square. (At the edge of the strata or sample area, the clusters of four points might not form squares due to irregular boundaries.)

Although the average contamination measure is computed in the usual manner as the sum of all observations divided by the number of them, the average may be considered as a weighted sum of the group means, where the weights are the number of observations in the group.

Figure 6.1 An Example of How to Group Sample Points from a Systematic Sample so that the Variance and Mean Can Be Calculated Using the Methodology for a Stratified Sample



The tests described in section 6.3 for simple random samples can be adapted for systematic samples by simply replacing the quantities $s/\sqrt{n}$ in equation (6.8) with the expression for the standard error given in equation (6.19). Box 6.18 gives a formula for an upper one-sided confidence interval for the true mean contamination when a systematic sample is treated as a stratified sample.

Box 6.17

Estimating the Mean, the Standard Error of the Mean, and Degrees of Freedom When a Systematic Sample Is Treated as a Stratified Sample

$$\bar{x} = \frac{1}{n} \sum_{h=1}^{L} n_h \bar{x}_h, \qquad (6.18)$$

where h denotes the h^{th} group, L is the total number of groups, n_h is the number of observations in group h, $\bar{x}_h$ is the mean of the observation in group h, and n is the total number of observations in the sample.

The estimated standard error of the mean, $s_{\bar{x}}$, can then be computed as:

$$s_{\bar{x}} = \sqrt{\frac{1}{n^2} \sum_{h=1}^{L} n_h s_h^2} \qquad (6.19)$$

where s_h^2 is the variance of the observations in group h as computed from the equations in section 6.2. The degrees of freedom are computed as: df = n-L.

Box 6.18

Formula for Upper One-sided Confidence Interval for the True Mean Contamination When a Systematic Sample Is Treated as a Stratified Sample

For example, the upper one-sided confidence interval for the true mean contamination is:

$$\mu_{U\alpha} = \bar{x} + t_{1-\alpha,df}\, s_{\bar{x}} \qquad (6.20)$$

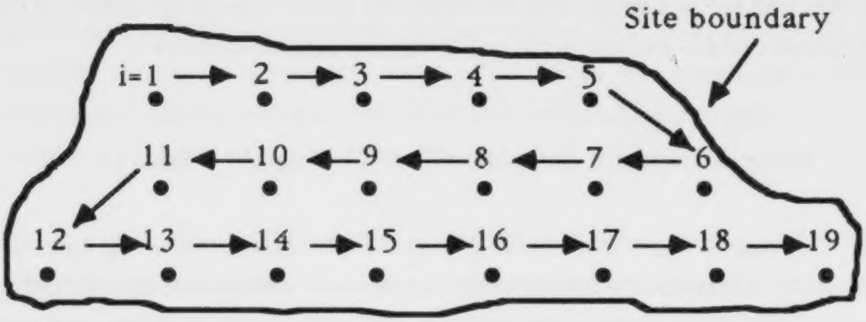
In this case, the sample area would be declared to be clean if $\mu_{U\alpha}$ is less than the cleanup standard; otherwise the sample area would be declared to be dirty.

6.5.2.3 Linearization and Estimates from Differences Between Adjacent Observations of a Systematic Sample

Another commonly used method is to linearize the systematic pattern by forming a serpentine association between each observation and the one preceding it in a serpentine pattern. Consider the example pattern in Figure 6.2. The numbers represent the sample points and their location in the sample area.

The numbers string the pattern into a linear sequence. The difference between the observations of an adjacent pair contain a systematic component that represents the "true" difference between them plus a random component. The systematic component represents bias but, since the two members of the pair are adjacent geographically, one would expect the systematic component of the difference to be small compared to, for example, comparing point 1 with point 19.

Figure 6.2 Example of a Serpentine Pattern



Numbers indicate the sequence (i) required for the calculations in Box 6.19.

To estimate the standard error from a serpentine pattern makes use of overlapping pairs. That is, point 1 is compared with point 2, 2 with 3, 3 with 4, and so on. The method gives a somewhat more precise estimate of the standard error. The method is shown in Box 6.19.

Box 6.19
Computational Formula for Estimating the Standard Error and Degrees of Freedom from Samples Analyzed in a Serpentine Pattern

$$s_{\bar{x}} = \sqrt{(1/2n) \sum_{i=2}^{n} (x_i - x_{i-1})^2/(n-1)} \qquad (6.21)$$

$$df \approx 2n/3$$

The associated number of degrees of freedom in Box 6.19 is given approximately by DuMouchel et al. (1973).

It should be noted that the serpentine pattern can be constructed by moving from top to bottom, from right to left, or diagonally within the systematic pattern. The pattern should be planned prior to sampling. If it is suspected that there will be a gradient in the data, say from top to bottom, then the serpentine pattern should be formed so that it follows the contours of the gradient to the extent that it is feasible to do so.

6.6 Using Composite Samples When Testing the Mean

"Compositing" refers to the process of physically combining and mixing several individual soil samples to form a single "composite" sample (see Rohde, 1976 and 1979; Duncan, 1962; Elder et al., 1980; Gilbert, 1987, Gilbert et al., 1989, and section 5.6.2 of this document). A primary advantage of compositing is that it reduces the number of lab analyses that must be performed.

Composite samples can be created using the following procedure:

- Collect the samples using a random or systematic sample design, collecting n soil samples from the field;
- Physically mix randomly selected groups of ten samples to create n/10 = m samples, which are sent to the lab for analysis; and

- Perform the statistical analysis on the m lab results to determine if the mean attains the cleanup standard.

In the procedure above, each soil sample sent to the lab was composited from 10 original samples; the compositing factor was 10. The compositing was accomplished by mixing the first 10 randomly selected samples, the second 10 randomly selected samples, etc. to get the final m samples to send to the lab. To specify how compositing is done, both the method of selecting the samples that get mixed together and the compositing factor must be specified. In addition, compositing requires that each original sample is the same or known physical size in terms of volume or weight and that the samples are very well mixed. These criteria may be difficult to achieve. This possible advantage will be reduced if the mixing is not complete or uses soil samples of different physical sizes. Nevertheless, as mentioned above, the number of lab analyses that must be performed may be greatly reduced.

Other considerations are the decisions related to how best to composite the original samples, the number of soil samples to collect, and the number of soil samples to send to the lab. If the laboratory error is large, compositing may provide little benefit. The specification of which samples to combine will be affected by the sample design and the variability across the sample area, among other things. For some types of soil or chemicals being tested, mixing will affect the laboratory analysis. For example, mixing samples with volatile organics may release contaminants.

Compositing can be a useful technique if the mean is to be tested, but must always be considered and implemented with caution. Compositing should never be used if percentiles or proportions are used as the attainment criteria. Other methods of compositing are discussed by Gilbert (1987). If compositing is considered, consultation with a statistician is recommended.

6.7 Summary

The methods in this chapter apply when the cleanup standard is intended to control the average conditions at the site, not simply the average of the sample. The mean estimated from a sample must be sufficiently below the cleanup standard to ensure with confidence that the entire site is below the cleanup standard.

CHAPTER 6: DETERMINING WHETHER THE MEAN CONCENTRATION OF THE SITE IS LESS THAN A CLEANUP STANDARD

Basic formulas are provided to calculate the mean, variance, or standard deviation for a sample of data. The standard deviation provides a measure of the variability of the sample data and is used to obtain estimates of standard errors and confidence limits. These statistics help determine how far the sample mean must be below the cleanup standard to ensure with reasonable confidence that the site mean is below the cleanup standard.

For a random sampling, the number of soil samples required depends on the anticipated variability of the soil measurements. To estimate the required sample size, some information about the standard deviation, σ, or the variance σ^2, is needed. Steps to estimate σ are discussed. Equations for determining sample size require the following quantities: cleanup standard (Cs), the mean concentration where the site should be declared clean with a high probability (μ_1), the false positive rate (α), the false negative rate (β), and the standard deviation (σ). The mean of the sampling data is an estimate of the mean contamination of the entire sample area. The use of an upper "confidence interval" provides an upper bound on the true sample area mean. When a one-sided 100 $(1-\alpha)$ percent upper confidence limit of the mean is less than the Cs, the site is judged clean.

Estimating the mean contamination from a stratified sample requires considerable advance knowledge about the relative costs and variability within each strata. Guidelines and formulae are given to assist in planning the sample sizes for a stratified sample and how many soil units should be collected in each stratum. They are also given for establishing the standard error, the approximate degrees of freedom for the standard error, and the upper one-sided confidence interval. If the upper one-sided confidence interval on the sample mean is below the cleanup standard (Cs), cleanup is verified.

If systematic sampling is used, special methods are required; these procedures are discussed and illustrated. To estimate the standard error for a systematic sample, formulae used for a simple random sample and a stratified sample may be applied, as can be the method of linearizing the systematic pattern into a serpentine pattern. Two estimates of the standard error are common when the points (sampling locations) have been linearized; these are discussed.

Compositing samples--the act of physically combining and mixing several individual soil samples to form a single composite sample--is discussed. Its primary advantage is that it reduces the number of lab analyses that must be performed.

7. DETERMINING WHETHER A PROPORTION OR PERCENTILE OF THE SITE IS LESS THAN A CLEANUP STANDARD

This chapter describes statistical procedures for determining with confidence whether a specified proportion of the soil is less than a cleanup standard. The extreme concentrations at a hazardous waste site are often of primary concern. In this case, an appropriate statistical test can be based on either a high percentile of the distribution of chemical measurements over the area, or on a large proportion of the area that has concentrations less than the cleanup standard. For example, the methods in this chapter apply if there is interest in verifying that a large percentage (e.g., 90, 95, or 99 percent) of the soil at the site has concentrations below the cleanup standard.

Throughout Chapter 7 the statistical evaluations are designed to detect when a large proportion of the site is less than a cleanup standard. However, there is another equivalent way of stating this objective: these evaluations are designed to ensure that no more than a small proportion of the site is above the cleanup standard. The numerical methods in this chapter are designed and presented in the context of the second approach. Therefore, we will be testing to verify that only a small proportion or percentage of the site (e.g., 10, 5, or 1 percent) exceeds the cleanup standard.

Two approaches to testing percentiles and proportions are discussed in this chapter:

- Exact and large sample nonparametric tests for proportions based on the binomial distribution; and
- A parametric test for percentiles based on tolerance intervals, which assumes the data have a normal distribution.

In the nonparametric approach, each soil sample measurement is designated as either equal to or above the cleanup standard, Cs and coded as "1," or below Cs and coded as "0." The analysis is based on the resulting data set of 0's and 1's. The proportion of the soil (or equivalently, the percentage of the area under investigation) at or above the cleanup standard can be estimated from the coded data. If the proportion of 1's is high, the site will be declared contaminated. On the other hand, if the proportion of 0's is sufficiently large, the area is considered to have attained an acceptable level of cleanup. A test based on proportions works with any concentration distribution and requires only that the cleanup standard be greater than the analytical method detection limit. However, this method has limitations because it does not consider how far above or below the Cs the data value is, only if it is above or below.

The second approach for testing percentiles of the concentration distribution, which does not require coding the data as above, is based on estimating a confidence interval for a percentile of the normal distribution. These intervals are called tolerance intervals (Guttman, 1970). The assumption that the data have a normal distribution (or that a suitable transformation of the data is approximately normal) is critical to this test. In addition, this method may be biased if more than 10 percent of the observations are below the detection limit.

The following sampling and analysis plans are discussed in the sections indicated:

- Simple random sampling for proportions (section 7.3);
- Stratified random sampling for proportions (section 7.5); and
- Simple random sampling for testing percentiles of a normal distribution (section 7.6).

7.1 Notation Used in This Chapter

The following notation is used throughout this chapter:

Cs	The cleanup standard relevant to the sample area and the contaminant being tested (see section 3.4 for more details).
P	The "true" but unknown proportion of the sample area with contaminant concentrations greater than the cleanup standard.
P_0	The criterion for defining whether the sample area is clean or dirty. According to the attainment objectives, the sample area attains the cleanup standard if the proportion of the sample area with contaminant concentrations greater than the cleanup standard is less than P_0, i.e., the sample area is clean if $P < P_0$.

H_0	The null hypothesis, which is assumed to be true in the absence of significant contradictory data. When testing proportions, the null hypothesis is that the sample area does not attain the cleanup standard; H_0: $P \geq P_0$.
α	The desired false positive rate for the statistical test to be used. The false positive rate for the statistical procedure is the probability that the sample area will be declared to be clean when it is actually dirty.
H_1	The alternative hypothesis, which is declared to be true only if the null hypothesis is shown to be false based on significant contradictory data. When testing proportions, the alternative hypothesis is that the sample area attains the cleanup standard; H_1: $P < P_0$.
P_1	The value of P under the alternative hypothesis for which a specified false negative rate is to be controlled. Think of P_1 as the value less than P_0 ($P_1 < P_0$) that designates a very clean area that must, with great certainty, be designated clean by the statistical test.
β	The false negative rate for the statistical procedure is the probability that the sample area will be declared to be dirty when it is actually clean and the true mean is P_1. The desired sample size n_d is selected so that the statistical procedure has a false negative rate of β at P_1. See section 2.1 and Table 2.1 for further discussion.
n_d	The desired sample size for the statistical calculations.
n	The final sample size, i.e., the number of data values available for statistical analysis including the concentrations that are below the detection level.
x_i	The contaminant concentration measured for soil sample i. For measurements reported as below detection, x_i = the detection limit.
y_i	The coded value of x_i. If the concentration in sample i is less than the cleanup standard ($x_i < Cs$), then $y_i = 0$. If the concentration in the sample is greater than or equal to the cleanup standard ($x_i \geq Cs$), then $y_i = 1$.

7.2 Steps to Correct for Laboratory Error

All of the procedures for estimating proportions and percentiles assume that the chemical concentrations can be measured with little or no error. If there is substantial variability in the measurement process, the corresponding estimates of proportions may be biased (Mee et al., 1986 and Schwartz, 1985). If an upper percentile (greater then the median) is being tested, the bias may be conservative. In other words, the sample area may be cleaner than the statistical test would indicate. This bias will be more important in some situations than in others:

- The measurement error causes no problems if the median (50th percentile) is tested;
- The measurement error is likely to be the greatest problem when the percentile to be tested is between the 75th and 99th; and
- The measurement error is likely to be the greatest problem if the true proportion of contaminated soil samples is close to the proportion being tested, i.e., the sample area just attains the cleanup standard.

There are three possible ways to reduce the bias:

- Use a more precise analytical method that has a smaller measurement error;
- Perform multiple laboratory measurements on each soil sample and use the average or median measurement in the statistical analyses (see the example in Box 7.1); or
- Perform more cleanup of the sample area than is required to attain the cleanup standard.

Box 7.1
Illustration of Multiple Measurement Procedure for Reducing Laboratory Error

Soil unit	Measurement (ppb) 1	2	3	Sample median (ppb)	Coded result
1	<50	95	101	95	0
2	75	105	102	102	1
3	<50	<50	55	<50	0
.	.	.	.	.	.
.	.	.	.	.	.

Detection limit - 50 ppb
Cs - 100 ppb

7.3 Methods for Simple Random Samples

This section describes statistical analysis procedures that apply when the criterion for deciding whether the site attains the cleanup standard is based on the proportion of contaminated soil units and when the soil samples are selected by simple random sampling. The basic steps involved in the data collection and analysis are:

- Determine the required sample size (section 7.3.1);
- Identify the locations within the site from which the soil units are to be collected, collect the physical samples, and send sampled material to laboratory for analysis (Chapter 5);
- Perform appropriate statistical analysis using the procedures described in sections 7.3.3 and 7.3.4 and, on the basis of the statistical analysis, decide whether the site requires additional cleanup.

Although the use of random samples is recommended, random sampling may not be practicable. An alternative is to select a systematic or grid sample using the procedures described in Chapter 5. Systematic samples may be easier to collect and will provide valid estimates of proportions, but may produce a poor estimate of sampling error.

7.3.1 Sample Size Determination

The sample sizes as computed in Box 7.2 are summarized in Tables A.7 through A.9 for selected values of P_0 and P_1 and for the following values of α and β: $\alpha = 0.01$, 0.05, and 0.10, and $\beta = 0.20$. In most cases, Tables A.7 - A.9 will be adequate for practical application. However, for values not in the tables, use equation (7.1) below. Notice that the cleanup standard is not required in order to determine the sample size.

Box 7.2
Computing the Sample Size When Testing a Proportion or Percentile

Given the quantities, P_0, P_1, α, and β, the sample size can be computed from the following formula:

$$n_d = \left\{ \frac{z_{1-\beta}\sqrt{P_1(1 - P_1)} + z_{1-\alpha}\sqrt{P_0(1 - P_0)}}{P_0 - P_1} \right\}^2 \qquad (7.1)$$

where $z_{1-\beta}$ and $z_{1-\alpha}$ are the critical values for the normal distribution with probabilities of $1-\alpha$ and $1-\beta$ (Table A.2).

7.3.2 Understanding Sample Size Requirements

To illustrate the use of the sample size tables, consider the following scenario (also see Box 7.3). A sample area will be considered clean if less than 20 percent of the area has concentrations of mercury greater than 1.5 ppm. That is, $P_0 = .20$ in this example. The null hypothesis is $H_0{:}P \geq .20$ and specifies that if 20 percent or more of the sample area has concentrations exceeding 1.5 ppm, the area is still considered dirty and requires further remedial action.

Further suppose that the site manager wants no more than a 5 percent chance of declaring the sample area to be clean when it is dirty (i.e., $\alpha = .05$). Moreover, the site manager wants to be 80 percent certain that if only 10 percent of the area has concentrations exceeding 1.5 ppm the site will be found clean. That is, for $P_1 = .10$, he wants the false negative rate to be moderately low, say 20 percent (i.e., $\beta = .20$). From Table A.8 (corresponding to values of $\alpha = .05$ and $\beta = .20$), the required sample size for $P_0 = .20$ and $P_1 = .10$ is $n_d = 83$.

It is evident from Tables A.7 - A.9 that as the value of P_1 approaches P_0, the required sample sizes become larger. For example, if the manager in the above example wanted the false negative rate to be 20 percent for $P_1 = .15$ (instead of $P_1 = .10$), the required sample size would be 368. Such a large sample size may be impractical for many waste site investigations. If the cleanup technology is designed to achieve levels that are only slightly less (P_1) than the cleanup objective (P_0), then many samples will be required to verify attainment of P_0. If the cleanup is highly effective and P_1 is well below P_0, then few samples will be required to verify cleanup.

Box 7.3
Example of How to Determine Sample Sizes
When Evaluating Cleanup Standards Relative to a Proportion

Soil has been removed from a lagoon bottom that previously contained corrosive waste. The exposed soil will be sampled to determine whether more excavation is required. Wanting to minimize the possibility of future health effects, the site will be judged in attainment of the cleanup standard if there is 90 percent confidence ($\alpha = .10$) that less than 10 percent ($P_0 = .10$) of the topsoil has concentrations exceeding the cleanup standard. The expected proportion of contaminated soil is low, less than 5 percent. The manager wants to be 80 percent confident ($\beta = .20$) that the sample area will be declared clean if the proportion of contaminated soil is less than 2 percent (P_1 = 2 percent).

From Table A.9, for $P_0 = 0.10$ and $P_1 = 0.02$, the required sample size is n = 39.

Using formula (7.1), from Table A.2, $z_{1-\alpha} = 1.282$ and $z_{1-\beta} = .842$ and:

$$n_d = \left\{ \frac{z_{1-\beta}\sqrt{P_1(1 - P_1)} + z_{1-\alpha}\sqrt{P_0(1 - P_0)}}{P_0 - P_1} \right\}^2$$

$$= \left\{ \frac{.842\sqrt{.02(.98)} + 1.282\sqrt{.10(.90)}}{.10 - .02} \right\}^2$$

$$= 39.4$$

7.3.3 Estimating the Proportion Contaminated and the Associated Standard Error

This section describes the computational procedures to be used to calculate the proportion contaminated (see Box 7.4) and related quantities necessary to evaluate attainment of the cleanup standard.

Box 7.4
Calculating the Proportion Contaminated and the Standard Error of the Proportion

Set $y_i = 1$ if the concentration in sample i is greater than the cleanup standard and $y_i = 0$ otherwise. If n = the total number of samples available for statistical analysis, the proportion of samples, p, above the cleanup standard can be calculated using the following equations:

$$r = \sum_{i=1}^{n} y_i \qquad (7.2)$$

where $y_i = 1$ if $x_i > Cs$ or $y_i = 0$ if $x_i \le Cs$

$$p = \frac{r}{n} \qquad (7.3)$$

The standard error, s_p, of the proportion p is

$$s_p = \sqrt{\frac{p(1 - p)}{n}} . \qquad (7.4)$$

These results are used to estimate upper one-sided confidence intervals, which allow determination of whether the site has attained the prescribed cleanup standard.

If the sample size is sufficiently large, an approximate confidence interval may be constructed using the normal approximation (see Box 7.5, section 7.3.4). If the sample size is small, an "exact" procedure should be used to calculate the confidence interval (see Box 7.6, section 7.3.5).

7.3.4 Inference: Deciding Whether a Specified Proportion of the Site is Less than a Cleanup Standard Using a Large Sample Normal Approximation

When $np \ge 10$ and $n(1-p) \ge 10$, the large sample normal approximation can be used for evaluating the statistical significance of the number of sample values equal to or above Cs. This condition will generally only be met for tests of percentiles between 10 and 90. If the condition is not met, the exact test should be used.

Box 7.5
Calculation of the Upper Confidence Limit on a Proportion
Using a Large Sample Normal Approximation

Compute the following:

$$P_U = p + z_{1-\alpha}\, s_p \qquad (7.5)$$

If $P_U < P_0$, conclude that the area has attained the cleanup standard.

If $P_U \geq P_0$, conclude that the area has not attained the cleanup standard.

7.3.5 Deciding Whether a Specified Proportion of the Site is Less Than the Cleanup Standard Using an Exact Test

If the normal approximation is not appropriate, the "exact" procedure described below should be used to test whether the proportion meets the cleanup standard. However, if the sample size is too small, it may not be possible to construct a useful decision rule with the stated false positive rate. These instances are indicated in the tables (Tables A.7 - A.9) used to perform the tests.

Use the following to perform the exact test:

- Given n, α, and P_0, determine the "critical value" of the test, $r_{\alpha;n}$, by referring to Table A.10. To use this table, α must be .01, .05 or .10, respectively. To determine the critical value, select the column for P_0 specified in the attainment objectives. Reading down the column find the first number greater than the sample size n. Move up one row and read $r_{\alpha;n}$, the critical value, in the leftmost column. If the number in the first row of the selected column is greater than the sample size, there are not enough data to perform the given test. If the bottom number in the selected column is less than the sample size, use the normal approximation above.
- From the sample, determine the number, r, of soil units that have chemical concentrations exceeding Cs. Compare r with $r_{\alpha;n}$.

- If $r < r_{\alpha;n}$, conclude that the area has attained the cleanup standard.
- If $r \geq r_{\alpha;n}$, conclude that the area has not attained the cleanup standard.

For values of n, α, and P_0 that are not given in the tables, the critical value for the "exact" test may be determined directly using the algorithm below or using an equivalent procedure from Brownlee (1965, p. 148-150) based on the F distribution.

Step 1 Compute $f(0) = (1 - P_0)^n$.

Step 1a If $f(0) \geq \alpha$, then set $r_{\alpha;n} = 0$ and stop. Note that if $f(0) > \alpha$, a test with the specified false positive rate is not possible; the actual false positive rate would be f(0).

Step 1b If $f(0) < \alpha$, go to Step 2.

Step 2 Compute

$$f(1) = n \left(\frac{P_0}{1 - P_0}\right) f(0) \qquad (7.6)$$

where f(0) is computed in Step 1.

Step 3 Next, compare f(0) + f(1) with α. If $f(0) + f(1) \geq \alpha$, set $r_{\alpha;n} = 0$, and stop. If $f(0) + f(1) < \alpha$, define a "temporary" variable, y, and set y = 1. Go to Step 4.

Step 4 For the given value stored in the temporary variable, y, compute f(y) using the recursion formula below:

$$f(y) = \left(\frac{n-y+1}{y}\right) \left(\frac{P_0}{1 - P_0}\right) f(y-1). \qquad (7.7)$$

Step 5 Compare $f(0) + f(1) + ... + f(y)$ with α. If $f(0) + f(1) + ... + f(r) \geq \alpha$, set $r_{\alpha;n} = y$ and stop. If $f(0) + f(1) + ... + f(r) < \alpha$, increment the temporary variable by 1, i.e., set y = y+1, and go to Step 3. Repeat Steps 4 and 5 until the process stops and $r_{\alpha;n}$ has been determined.

Box 7.6 gives an example of an inference based on the "exact" test described above.

Box 7.6
An Example of Inference Based on the Exact Test

Assume that only 9 samples collected from 203 sample locations have concentrations greater than the cleanup standard, i.e., $r = 9$, and remember that $n = 191$, $\alpha = .05$, and $P_0 = .05$.

Using Table A.10 read down the column headed by $P_0 = .05$ and find the first number greater than the sample size, in this case 208 in row 6. Go up one row and read $r_{\alpha:n}$ from the lefthand column. The value in the left column and fifth row is $4 = r_{\alpha:n}$.

Since $r > 4$, the sample area does not attain the cleanup standard.

7.4 A Simple Exceedance Rule Method for Determining Whether a Site Attains the Cleanup Standard

One of the most straightforward applications of the methods in this chapter involves the design of zero or few exceedance rules. To apply this method, simply require that a number of samples be acquired and that zero or a small number of the concentration measurements be allowed to exceed the cleanup standard. This kind of rule is easy to implement and evaluate once the data are collected; it only requires specification of the sample size and number of exceedances as indicated in Table 7.1.

In addition, these rules also have statistical properties. For example, the more samples collected, the more likely that one sample will exceed a cleanup standard. That is, it is more likely to measure a rare high value with a larger sample. In addition, the larger the proportion of the site that must have concentrations below the cleanup standard, the more soil samples that will be required to document this with certainty. Finally, because of the chance of outliers, it may be that the rule that allows one or more exceedances would be preferred in order to still have the site judged in attainment of the cleanup standard. If more exceedances are allowed, more soil samples are required to maintain the same statistical performance and proportion of the site that is clean.

Table 7.1 illustrates these tendencies and offers selected sample sizes and exceedance rules as a function of statistical performance criteria. For example, if there is interest in: verifying that 99 percent of the site is below a cleanup standard; keeping the chance of saying the site is clean when it is dirty at 1 percent; and allowing no exceedances of the cleanup standard, then 459 soil samples would be required. If 459 samples were obtained and none of them exceeded the cleanup standard, there is 99 percent confidence that 99 percent of the site is less than the cleanup standard. If three exceedances were allowed and the same statistical performance criteria were required then 1001 soil samples would be required and 998 of the measurements would have to be less than the cleanup standard.

On the other hand, if the statistical performance criteria are relaxed, sample size requirements decrease. For example, if there is interest in allowing no exceedances and a false positive rate of 90 percent that 90 percent of the site is less than the cleanup standard, then 22 samples would have to be obtained and all results would have to be less than the cleanup standard. If three exceedances were permitted and the same statistical criteria were applied, then 65 samples would be required and 62 of the measurements would have to be less than the cleanup standard.

7.5 Methods for Stratified Samples

In some circumstances it may be useful to establish a stratified sampling regime as discussed in Chapter 5. If the waste can be divided into homogeneous subareas, the precision of an estimated proportion can often be improved through the use of a stratified sample. These homogeneous areas from which separate samples are drawn are referred to as "strata," and the combined sample from all areas is referred to as a "stratified sample."

The statistical procedures discussed here apply when the criterion for deciding whether the site attains the cleanup standard is based on the proportion of contaminated soil units. The basic steps involved in the data collection and analysis are:

- Determine the required sample sizes for each stratum using the equation in section 7.5.1;

- Within each stratum, identify the locations within the site from which the soil units are to be selected, collect the physical samples, and send sampled material to laboratory for analysis (Chapter 5); and
- Perform appropriate statistical analysis using the procedures that follow (sections 7.5.2 and 7.5.3) and, on the basis of the statistical results, decide whether the site has attained the cleanup standard.

The tests described in this section assume that soil samples within each stratum are collected randomly. Although the use of simple random samples is recommended, simple random sampling may not always be practicable. An alternative method would be to select a systematic (grid) sample; however, this type of sampling should be approached with caution as described in section 7.3 and Chapter 6.

7.5.1 Sample Size Determination

Determination of the appropriate sample size is complicated in stratified sampling because there are many ways the sample can be allocated to strata. For example, if 100 soil units will be sampled, a decision must be made on whether to allocate the sample equally among strata, in proportion to the relative size of each strata, or according to some rules. There are methods for determining the "optimum" allocation; however, these require considerable advance knowledge about the underlying variability of each strata. Consequently, the equations below are general guidelines to assist in planning the sample sizes for a stratified sample. These guidelines will cover many field situations. For more complex situations, a text such as Cochran (1977) should be consulted.

The formulas for sample size use the following notation, where h indicates the stratum number:

h	As a subscript, indicates a value for a stratum within the sample area rather than for the entire sample area.
n_{hd}	The desired sample size for the statistical calculations.
n_h	The final sample size, the number of data values available for statistical analysis including the concentrations that are below the detection level.
W_h	Proportion of the volume of soil in the sample area which is in stratum h

C_h	Cost of collecting, processing, and analyzing one additional soil sample, on a relative scale.
L	The number of strata.
y_{hi}	The scored concentration data, where $y_{hi} = 1$ if the measured concentration is greater than the cleanup standard and 0 otherwise.

Table 7.1 Selected information from Tables A.7 - A.9 that can be used to determine the sample sizes required for zero or few exceedance rules associated with various levels of statistical performance and degrees of cleanup

Chance of Saying the Site is Clean When It is Dirty (Certainty)	Proportion of the Site That Is Clean	Sample Size Requirements Under Various Numbers of Allowed Exceedances of the Cleanup Standard			
False Positive Rate, Alpha (1 - Alpha)	$1-P_0$	Number of Allowed Exceedances			
		0	1	3	5
.01	.99	459	662	1001	1307
(.99)	.95	90	130	198	259
	.90	44	64	97	127
.05	.99	299	473	773	1049
(.95)	.95	59	93	153	208
	.90	29	46	76	103
.10	.99	230	388	667	926
(.90)	.95	45	77	132	184
	.90	22	38	65	91

Once the sample area has been divided into strata, it is necessary to decide how many soil units should be collected in each stratum. The equations below will provide an "optimal" sample size for each stratum provided that the following information is available:

- The physical size of the stratum;
- The cost of sampling and processing a soil unit selected from the stratum;
- The underlying proportion of the soil units in the stratum that are contaminated, i.e., have chemical concentrations exceeding the specified cutoff, Cs; and
- The overall desired accuracy of the test.

An optimum sample allocation to each stratum will produce the most accurate measure of the proportion of soil contaminated across strata in the entire sample area for a fixed total cost. In what follows, n_h will denote the corresponding sample size to be selected from stratum h. Thus, the total sample size n, is calculated as follows: $n = n_1+n_2+...+n_L$.

Although the sample size equations assume that the quantities C_h and P_h are known, reasonable assumptions can be used, following the rules below (see Box 7.7):

- If the relative sampling costs, C_h, are not known or all strata are assumed to have the same cost for an additional sample, set $C_h = 1$ for all strata;
- If data are not available to provide an estimate of P_h in some strata, set $P_h = P_o$ for those strata.

The expected number of contaminated samples in stratum h is $P_h * n_{hd}$. It is recommended that the expected number of contaminated samples in each stratum be at least 5 for calculation of reliable confidence intervals. Occasionally this may require increasing the sample size in one or more strata.

Box 7.7
Computing the Sample Size for Stratum h

Given C_h, P_h, and W_h, the sample size for stratum h should be computed as:

$$n_{hd} = P_h(1 - P_h) * \left\{\sum_{h=1}^{L} W_h \sqrt{C_h}\right\} * \left\{\frac{z_{1-\alpha} + z_{1-\beta}}{P_0 - P_1}\right\}^2 * \frac{W_h}{\sqrt{C_h}} \qquad (7.8)$$

7.5.2 Calculation of Basic Statistics

This section describes the computational procedures to be used to calculate the quantities necessary to evaluate attainment of the cleanup standard on the basis of the overall proportion of contaminated samples. Box 7.8 gives sample size calculations for stratified sampling.

Use the formula below in Box 7.9 for calculating an overall proportion of exceedance from a stratified sample. Note that the overall sample proportion, denoted by p_{st}, is simply a weighted average of the individual stratum means.

Box 7.8
Sample Size Calculations for Stratified Sampling

At a site with heavy metal contamination, the sample area has been divided into two strata, one consisting of high elevation areas, another of low elevation areas which received most of the historical runoff. The strata are the same volume ($W_1 = .5,\ W_2 = .5$) The expected proportion of contaminated soil is 5 percent on the higher ground and 10 percent in the lower area ($P_1 = .05,\ P_2 = .10$). Due to difficult access and low trafficability in the lower area, the cost of sampling is twice what it is on the high ground ($C_1 = 1,\ C_2 = 2$). EPA has decided that less than 10 percent of the soil can have concentrations over the cleanup standard (with a confidence of 90 percent, $\alpha = .10$). The site manager must be able to conclude that the site is clean with a confidence of 80 percent ($\beta = .20$) at an overall contamination proportion of 4 percent.

To determine the sample size, the site manager first determines:

$$z_{1-\alpha} = 1.282,\quad z_{1-\beta} = .842$$

from Appendix A. Then, following equation (7.8):

$$\left\{\frac{z_{1-\alpha} + z_{1-\beta}}{P_0 - P_1}\right\}^2 = \left\{\frac{1.282 + .842}{.10 - .04}\right\}^2 = 1,253$$

$$\left\{\sum_{h=1}^{L} W_h\sqrt{C_h}\right\} = \left\{(.5 * \sqrt{1}) + (.5 * \sqrt{2})\right\} = 1.207$$

$$n_{hd} = P_h(1 - P_h) * 1.207 * 1,253 * \left\{\frac{W_h}{\sqrt{C_h}}\right\}$$

and

$$n_{1d} = .05(1 - .05) * 1.207 * 1,253 * \frac{.5}{\sqrt{1}} = 35.9$$

$$n_{2d} = .10(1 - .10) * 1.207 * 1,253 * \frac{.5}{\sqrt{2}} = 48.1$$

Rounding up, the samples sizes of the strata are:

$$n_{1f} = 36, \text{ and } n_{2f} = 48$$

Box 7.9
Calculating an Overall Proportion of Exceedances and the Standard Error of the Proportion From a Stratified Sample

$$p_h = \frac{\sum_{i=1}^{n} y_{hi}}{n_h} \qquad (7.9)$$

p_h = the sample proportion of units in stratum h that have chemical concentrations exceeding Cs.

The estimated **overall proportion** of soil units that have chemical concentrations exceeding Cs is given by the formula below:

$$p_{st} = \sum_{h=1}^{L} W_h \, p_h \qquad (7.10)$$

Use equation (7.11) to estimate the **standard error** of p_{st}. The standard error is required for constructing an approximate decision rule and also for establishing confidence limits around the actual population proportion.

$$s_{p_{st}} = \sqrt{\sum_{h=1}^{L} W_h^2 \, \frac{p_h(1 - p_h)}{n_h}} \qquad (7.11)$$

In Box 7.10, use the equation (7.12) to compute the **upper limit of the one-sided confidence interval.**

Box 7.10
Calculating the Upper Limit of the One-sided
Confidence Interval on an Estimate of the Proportion

$$p_{U\alpha} = p_{st} + z_{1-\alpha} s_{p_{st}} \qquad (7.12)$$

where p_{st} is the computed overall proportion of contaminated units, and $s_{p_{st}}$ is the corresponding standard error. The value of $z_{1-\alpha}$ can be obtained from Table A.2.

The value $p_{U\alpha}$ designates an upper 100(1-α) percent one-sided confidence limit for the population proportion.

7.5.3 Inference: Deciding Whether the Site Meets Cleanup Standards

The upper one-sided confidence limit, $P_{U\alpha}$, is used for testing the hypothesis that $1 - P_0$ of the site attains the specified cleanup standard. Use the following rules to decide whether or not the site attains the cleanup standard:

If $p_{U\alpha} < P_0$, conclude that the site meets the cleanup standard.

If $p_{U\alpha} \geq P_0$, conclude that the site does not meet the cleanup standard.

See Box 7.11 for an example of an inference for proportions using stratified sampling.

Box 7.11
Inference for Proportions Using Stratified Sampling

Following the example in Box 7.8, all 434 samples from stratum 2 were collected; however, of the 324 samples in stratum 1, four were lost due to a lab error, leaving 320 samples for the analysis. The proportion of samples collected in each strata that had concentrations greater than or equal to the cleanup standard are: .0531 in strata 1 (the higher ground) and .0922 in strata 2.

Using equation (7.10)

$$p_{st} = \sum_{h=1}^{L} W_h \, p_h = .5 * .0531 + .5 * .0922 = .0727$$

Using equation (7.11)

$$s_{p_{st}} = \sqrt{\sum_{h=1}^{L} W_h{}^2 \frac{p_h(1 - p_h)}{n_h}}$$

$$= \sqrt{\frac{.25 * .0531(1 - .0531)}{320} + \frac{.25 * .0922(1 - .0922)}{434}} = .0094$$

Using equation (7.12)

$$p_{U\alpha} = p_{st} + z_{1-\alpha} \, s_{p_{st}} = .0727 + 1.282 * .0094 = .0848$$

Since .0848 is less than P_0 (.10), based on the proportion of contaminated samples, the sample area attains the cleanup standard.

7.6 Testing Percentiles from a Normal or Lognormal Population Using Tolerance Intervals

Tolerance intervals assume that the distribution of concentration measurements follows a normal distribution. Tolerance interval techniques are sensitive to the assumption that the data follow a normal distribution. This procedure is not robust to departures from the normality assumption.

If it is suspected that the data do not approximately follow a normal distribution, then either:

- Do not use the tolerance interval procedure and instead use the nonparametric procedures described in section 7.4; or
- Transform the data so that the transformed data more nearly approximate a normal distribution.

An approach that may be used to evaluate the assumption that the data follows a normal distribution is discussed in section 7.6.2. If the data are not normal and a transformation is being used then the transformation should be applied in the following manner. First, transform both the data and the cleanup standard. Then calculate the upper confidence limit on the percentile estimate of the transformed data. Compare the transformed upper limit with the transformed cleanup standard. Do not reverse; transform the upper confidence limit on the percentile for comparison with the untransformed cleanup standard. If stratified random sampling is used then consult Mee (1989).

7.6.1 Sample Size Determination

To determine the required sample size, the following terms need to be defined, P_0, P_1, α, β. Once these terms have been established, the following are obtained from Table A.2 and the equation in Box 7.12 is used to estimate the sample sizes:

$z_{1-\beta}$	the upper β-percentage point of a z distribution;
$z_{1-\alpha}$	the upper α-percentage point of a z distribution;
z_{1-P_0}	the upper P_0-percentage point of a z distribution; and
z_{1-P_1}	the upper P_1-percentage point of a z distribution.

Box 7.12
Calculating the Sample Size Requirements for Tolerance Intervals
(Guttman, 1970)

$$n_d = \left\{ \frac{z_{1-\beta} + z_{1-\alpha}}{z_{1-P_0} - z_{1-P_1}} \right\}^2 \qquad (7.13)$$

This sample size equation (7.13) requires smaller sample sizes than the corresponding formula in section 7.3.1. This happens because the tolerance intervals gain efficiency over the other methods in this chapter from the assumption that the data follow a normal distribution.

If the normal distribution is not followed, even after transformation, the procedure in this section is inappropriate. However, distributional form will not be evaluated until after the sample is collected and the data analyzed. At this point it may be decided to use the nonparametric procedures presented earlier in this chapter, but the sample size may not be sufficient to ensure the desired false negative rate and, therefore, may not be as sensitive as required.

Two example sample size calculations for tolerance intervals are shown in Box 7.13. The reduction in the required sample size between the nonparametric test and the tolerance interval test can be compared. The comparable sample sizes for the nonparametric test are 1990 samples for example #1 and 315 samples for example #2. In both examples, the tolerance interval method requires fewer samples, provided that it can be reasonably concluded that the data follow a normal distribution.

7.6.2 Testing the Assumption of Normality

The statistical tests used for evaluating whether or not the data follow a specified distribution are called goodness-of-fit tests. There are many different tests and references demonstrating the evaluation of normality (e.g., Conover, 1980; D'Agostino, 1970; Filliben, 1975; Mage, 1982; and Shapiro and Wilk, 1965). If a choice is available, the Shapiro-Wilk or the Kolmogorov-Smirnov test with Lilliefors critical values are suggested. For easy application, Geary's test described by D'Agostino (1970) can be used.

Box 7.13
Calculating Sample Size for Tolerance Intervals--Two Examples

Following are two examples of the computation required to calculate the sample size when testing percentiles using confidence intervals.

Example #1

$P_0=.010 \quad z_{1-P_0} = z_{.990} = 2.326$

$P_1=.005 \quad z_{1-P_1} = z_{.955} = 2.576$

$\alpha=.05 \quad z_{1-\alpha} = z_{.950} = 1.645$

$\beta=.2 \quad z_{1-\beta} = z_{.800} = 0.842$

$$n_d = \left\{\frac{.842 + 1.645}{2.326 - 2.576}\right\}^2 = \left\{\frac{2.487}{-.250}\right\}^2 = 98.96 \approx 99$$

Example #2

$P_0=.10 \quad z_{1-P_0} = z_{.90} = 1.282$

$P_1=.05 \quad z_{1-P_1} = z_{.95} = 1.645$

$\alpha=.05 \quad z_{1-\alpha} = z_{.95} = 1.645$

$\beta=.05 \quad z_{1-\beta} = z_{.95} = 1.645$

$$n_d = \left\{\frac{1.645 + 1.645}{1.282 - 1.645}\right\}^2 = \left\{\frac{3.29}{-.363}\right\}^2 = 82.14 \approx 82$$

7.6.3 Inference: Deciding Whether the Site Meets Cleanup Standards Using Tolerance Limits

The test of significance will be performed by estimating the upper confidence interval on the point below which at least $(1-P_0)*100$ percent of the data falls: the $[(1-P_0)*100]$th percentile. For example, the concentration measurement associated with $P_0 = .05$ is the value below which at least 95 percent of the data falls. The concentration measurement associated with $P_0 = .05$ will be calculated from the sample mean and standard deviation, $\bar{x}$ and s, as well as the constant k. The constant, k, necessary for finding the upper tolerance limit, T_u is found using values of α, P_0, n, and T in Table A.3. For values of k not shown in Table A.3, see Guttman (1970). With these three quantities an estimated upper tolerance limit will be calculated for the desired percentile using the equation in Box 7.14.

Box 7.14
Calculating the Upper Tolerance Limit

$$T_u = \bar{x} + ks. \tag{7.14}$$

If T_u is greater than the cleanup standard, then it is concluded that the site fails to meet the cleanup standard.

Box 7.15 presents data and calculations that illustrate use of tolerance intervals to test for percentiles with lognormal data.

Box 7.15
Tolerance Intervals: Testing for the 95th Percentile with Lognormal Data

The following data were collected to determine if the 95th percentile of the concentrations was below the cleanup standard of 100 ppm (with a false positive rate of 1 percent). The data is assumed to follow a lognormal distribution, therefore logarithm of the data (the transformed data) are analyzed. In the following, x refers to the original data and y refers to the transformed data. Because the log of the data is used, the upper confidence interval on the 95th percentile of the data must be compared to the log of the cleanup standard (ln(100)=4.605). Twenty samples were obtained.

x	ln(x)=y	y^2
34	3.526	12.433
79	4.369	19.088
38	3.638	13.235
62	4.127	17.032
6	1.792	3.211
14	2.639	6.964
20	2.996	8.976
31	3.434	11.792
42	3.738	13.973
36	3.584	12.845
57	4.043	16.346
24	3.178	10.100
57	4.043	16.346
188	5.236	27.416
26	3.258	10.615
45	3.807	14.493
46	3.829	14.661
83	4.419	19.528
25	3.219	10.362
33	3.497	12.229
Total	72.372	271.645

Using the logarithms as the data to analyze, the sample mean is:

$$\bar{y} = \frac{72.372}{20} = 3.619$$

The standard deviation, s, can be calculated using equations (6.2) and (6.3):

$$s^2 = \frac{271.645 - 20(3.619)}{19} = .511 \qquad s = \sqrt{.511} = 0.715$$

For a sample size of 20, $\alpha = .01$ and $P_0 = 5$ percent, $k = 2.808$ (from Table A.5). Finally, T_u can be calculated using equation (7.14):

$$T_u = 3.619 + 2.808(.715) = 5.627$$

Since 5.627 is greater than 4.605 (the cleanup standard in logged units), the sample area does not attain the cleanup standard.

7.7 Summary

These methods can apply to the 50th percentile or median as an alternative to the mean or to a high percentile such as the 90th, 95th, or 99th. High percentile criteria apply when the clean up standard is viewed as a value that should be rarely exceeded at the site. Similar to testing the mean, the proportion of soil samples above the cleanup standard must be sufficiently low to ensure with confidence that the proportion of soil at the site meets the established percentile.

Two approaches to testing whether proportions or percentiles of the soil at a site are less than the cleanup standard are discussed:

- Exact and large sample nonparametric tests for proportions based on the binomial distribution; and
- A parametric test for percentiles based on tolerance intervals, which assumes the data have a normal distribution.

The first approach, or test, works with any concentration distribution and requires only that the cleanup standard(s) be greater than the analytical method detection limit. For testing proportions, simple random and stratified random sampling are discussed.

All of the procedures discussed assume that the chemical concentrations can be measured with little or no error; variability in measurement may bias the corresponding estimates of proportions. Ways to reduce the potential bias are discussed.

For simple random samples, the basic steps involved and that are discussed include the following:

- Determine the required sample size;
- Identify locations within the site from which soil units are to be collected, collect the samples, and send them to the laboratory; and
- Perform the statistical procedures described in this chapter, and then decide whether the site needs additional cleanup.

CHAPTER 7: DETERMINING WHETHER A PROPORTION OR PERCENTILE OF THE SITE IS LESS THAN A CLEANUP STANDARD

The implementation of simple exceedance rules in the statistical plan design requires that a certain number of samples be acquired and that none or a few of the concentration measurements be allowed to exceed the cleanup standard. The more exceedances allowed, the more soil samples that need to be collected to maintain the statistical performance and proportion of the site that is clean. Sample sizes and exceedance rules as a function of statistical performance criteria are presented in the chapter.

If stratified sampling is chosen, the basic steps involved include the following:

- Determine the required sample sizes for each stratum;
- Within each stratum, identify the locations within the site from which the soil units are to be collected, collect the samples, and send them to the laboratory; and
- Perform the statistical procedures described in this chapter, and then decide whether the site needs additional cleanup.

The use of tolerance intervals, which is discussed next in this chapter, assumes that the distribution of concentration measurements follows a normal distribution. Techniques for using tolerance intervals, including the transformation of lognormal data to a normal distribution, are included with two examples of sample size calculation and other relevant equations.

8. TESTING PERCENTILES AND PROPORTIONS USING SEQUENTIAL SAMPLING

This chapter discusses sequential sampling as a method for testing percentiles. With sequential sampling, a statistical test is performed after each sample or small batch of samples is collected and analyzed. The statistical test determines whether an additional sample should be collected or whether the sample area is judged to have or have not attained the cleanup standard.

Chapters 6 and 7 dealt with statistical tests that are based on samples of a predetermined size. Fixed sample size methods will sometimes require that an unnecessarily large sample size be used in order to meet the stated precision requirements. This can be avoided by using a sequential procedure. Sequential procedures terminate when enough evidence is obtained to either accept or reject the null hypothesis, and thus, sequential tests can respond quickly to very clean or very contaminated sites. Sequential procedures will also yield a lower sample size on the average than the fixed sample size procedure even when the true level of P is not greatly different from P_0.

Decisions based on sequential sampling methods will be particularly useful in conjunction with the "rapid turnaround" analytical methodologies that are being used more often at Superfund sites. Devices that measure volatile soil gases, H-NU's, ion specific probes, or onsite scanning laboratories can be used much more rapidly and extensively than conventional intensive laboratory extraction, identification, and quantification methods. Without rapid turnaround and the potential for additional sampling within a day or two, sequential methods are not useful because of the cost to remobilize a sampling team and the time required for laboratory processing. Nevertheless, "rapid turnaround" technology is typically less accurate than conventional methods and therefore, despite the larger sample sizes that are possible, should be applied in an orderly and thoughtful manner.

References on sequential analysis include: Armitage (1947), Wetherill (1975), Siegmund (1985), Sirjaev (1973), and Wald (1973).

8.1 Notation Used in This Chapter

The following notation is used throughout this chapter:

Cs	The cleanup standard relevant to the sample area and the contaminant being tested (see section 3.4 for more details).
P	The "true" but unknown proportion of the sample area with contaminant concentrations greater then the cleanup standard.
P_0	The criterion for defining whether the sample area is clean or dirty. According to the attainment objectives, the sample area attains the cleanup standard if the proportion of the sample area with contaminant concentrations greater than the cleanup standard is less than P_0, i.e., the sample area is clean if $P < P_0$.
H_0	The null hypothesis, which is assumed to be true in the absence of significant contradictory data. When testing proportions, the null hypothesis is that the sample area does not attain the cleanup standard; H_0: $P \geq P_0$.
α	The desired false positive rate for the statistical test to be used. The false positive rate for the statistical procedure is the probability that the sample area will be declared to be clean when it is actually dirty.
H_1	The alternative hypothesis, which is declared to be true only if the null hypothesis is shown to be false based on significant contradictory data. When testing proportions, the alternative hypothesis is that the sample area attains the cleanup standard; H_1: $P < P_0$.
P_1	The value of P under the alternative hypothesis for which a specified false negative rate is to be controlled.
β	The false negative rate for the statistical procedure is the probability that the sample area will be declared to be dirty when it is actually clean and the true mean is P_1. The desired sample size n_d is selected so that the statistical procedure has a false negative rate of β.
k	The cumulative number of soil units that exceed the cleanup standard Cs.
n	The number of soil units evaluated.

8.2 Description of the Sequential Procedure

In the sequential testing procedures developed by Wald (1973), sampling is performed by analyzing one soil unit at a time until enough data have been collected to either reject the null hypothesis in favor of the alternative hypothesis, or accept the null hypothesis.[1] The expected sample size, using this sequential procedure, will be approximately 30 to 60 percent lower than the corresponding fixed sample size test with the same α, β, P_0, and P_1. The sequential procedure will be especially helpful in situations where contamination at the site is very high or very low. In these situations the sequential procedure will quickly accumulate enough evidence to conclude that the site either fails to meet or meets the cleanup standard. However, it must be emphasized that the actual sample size of the sequential procedure for a given site could be larger than for the fixed sample size methods (see section 8.3).

Wald's sequential procedure consists of forming an acceptance and rejection region for the cumulative number of contaminated soil units relative to the total number of soil units evaluated. Figure 8.1 shows graphically how the procedure operates. The horizontal axis, denoted by n, represents the number of soil units evaluated. The vertical axis represents the cumulative number of contaminated soil units after n soil unit evaluations. The two lines in the graph establish the boundaries of the acceptance and rejection regions for the test. The intersection of these lines, C_A and C_B, with the vertical axis and their slope, are important parameters of this sequential procedure.

The sampled soil units are evaluated one at a time, and after each evaluation, the cumulative number or sum of contaminated units (i.e., soil units with concentrations exceeding the cleanup standard, Cs) is determined. If the cumulative sum crosses the topmost line into the acceptance region, the hypothesis of contamination is accepted. If the cumulative sum stays low and enters the rejection region below the second (lowermost) line, it is concluded that the site is not contaminated (i.e., the null hypothesis of contamination is rejected). Otherwise, the process continues; that is, another soil unit is evaluated, and the new cumulative sum is compared with the boundary values to determine whether to accept or reject the null hypothesis or to continue evaluating soil units. In

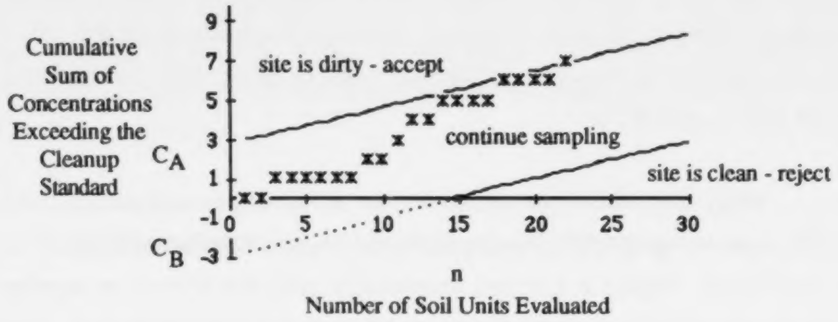
---

[1] The procedure in Wald's book is for a test of $P_1 > P_0$. In the present situation this has been reversed. To adapt the sequential procedure to this situation, the roles of α and β were reversed. The corresponding acceptance and rejection regions of the graphs were also reversed.

Figure 8.1 the process terminates after 22 soil units have been evaluated, at which time the null hypothesis that the area is contaminated is accepted.

Note that several soil samples can be collected and analyzed at the beginning of the sequential process, since some minimum number of results must be available before a decision can be reached.

Figure 8.1 Graphic Example of Sequential Testing



8.3 Sampling Considerations in Sequential Testing

It may be impractical to randomly collect a soil unit, chemically analyze the soil unit, and then decide whether or not to acquire the next unit. Instead, multiple soil units can be selected initially using the simple random sampling procedures described in section 5.2. The sampled soil units can then be chemically analyzed and each result evaluated individually in random order, until the sequential procedure terminates. It may also be possible, provided that the holding times or other analytical criteria are not violated, to chemically analyze samples one at a time.

In situations where contaminant concentrations at the site are marginally different from the cleanup standard, the sequential procedure can be expected to require more samples until the sample size approaches the sample size required for the fixed sample procedure. However, this is only an expectation, so in some situations where the actual contamination is close to the cleanup standard, the sequential procedure can require a substantially larger sample than the fixed sample procedure. In this situation, a cutoff rule is suggested. If the sequential procedure requires a sample size twice the sample size required for the fixed procedure, then the sequential sampling should be stopped and a decision made on the data collected up to that point. Procedures for accommodating this situation are discussed in Box 8.2.

Also, as with all of the procedures in this manual, the site is assumed to be at steady state during sampling. During the sequence of sampling the soil concentrations should not be changing. Sequential sampling and analysis does not imply that changes over time are being evaluated or that the progress of cleanup is being monitored. Sequential sampling is performed during steady state conditions, only to reduce the sample size required for a decision.

8.4 Computational Aspects of Sequential Testing

As was the case for the fixed sample tests described in earlier chapters, the following quantities must be defined to implement the sequential testing procedure: Cs, P_0, P_1, α, and β. Box 8.1 describes the method for establishing the acceptance and rejection boundaries described in Figure 8.1.

Denote the Qth percentile of chemical concentrations by X_Q. To test whether $X_Q \geq Cs$ or greater (i.e., the site fails to meet the cleanup standard) against the hypothesis that $X_Q < Cs$ (the site meets cleanup standards), set $P_0 = 1 - Q$, and set the maximum allowable error rate for falsely rejecting that the true percentile is Cs (i.e., false positive rate) to α. If the Qth percentile is really less than Cs (indicating that fewer than P_0 of the area is contaminated), specify the minimum value of this percentile, $P_1 < P_0$, that should be detected with at least a probability of $1 - \beta$.

To test whether the Qth percentile is equal to Cs, the sequential procedure is formatted by calculating the sequential procedure acceptance and rejection criteria as described in Box 8.1. Then follow the steps in Box 8.2 to decide whether the site attains the cleanup standard.

Box 8.1
Defining Acceptance and Rejection Criteria for the Sequential Tests of Proportions

Let ln(x) denote the natural logarithm of x. Given α, β, P_0, and P_1, compute:

1) $B = \ln(\frac{\alpha}{1-\beta})$ and $A = \ln(\frac{1-\alpha}{\beta})$;

2) $R_1 = (\frac{1 - P_0}{1 - P_1})$ and $R_2 = \frac{P_0}{P_1}$;

3) Use these values computed in (1) and (2) to determine the slope of the two lines defining the rejection and acceptance regions,

$$M = \frac{\ln(R_1)}{\ln(\frac{R_1}{R_2})} ;$$

and the points at which the two lines cross the vertical axis,

4) $C_A = \frac{A}{\ln(\frac{R_2}{R_1})}$ and $C_B = \frac{B}{\ln(\frac{R_2}{R_1})}$.

5) Finally, compute the desired sample size for the corresponding fixed sample size procedure,

$$n_d = \{\frac{z_{1-\beta}\sqrt{P_1(1-P_1)} + z_{1-\alpha}\sqrt{P_0(1-P_0)}}{P_0 - P_1}\}^2 \qquad (8.1)$$

8.5 Inference: Deciding Whether the Site Meets Cleanup Standards

Box 8.2
Deciding When the Site Attains the Cleanup Standard

1) Calculate the sequential procedure acceptance and rejection criteria described in Box 8.1.

2) After each evaluation calculate the cumulative number of soil units that exceed the cleanup standard, Cs:

$$k = \sum_{i=1}^{n} y_i , \qquad (8.2)$$

where $y_i = 1$ if the i-th sample was above the cleanup standard, and $y_i = 0$ otherwise; and where n is the number of soil units evaluated up to this point. Compare the current value of k against the current critical value to decide whether to accept or reject the null hypothesis or to continue sampling.

3) Starting with $n = 1$, if $k \geq nM + C_A$, then stop evaluating samples and accept H_0: $P \geq P_0$. Conclude that the site is dirty and requires additional cleanup.

4) If $k \leq nM + C_B$, then stop evaluating samples and reject H_0 in favor of $P \leq P_1$. Conclude that the site is clean.

5) If neither of the two conditions above is met, continue sampling and evaluation.

6) If the number of soil units that has currently been evaluated exceeds $2.0^* n_f$, stop the sampling and:

accept H_0: $P \geq P_0$ if $k \geq nM + \frac{C_A + C_B}{2}$ or

accept H1: $P \leq P_1$ if $k < nM + \frac{C_A + C_B}{2}$.

Rule 6 provides an approximate test and will have only a small effect on the actual levels of α and β (see Wald, 1973).

If the conclusion in step 3 is reached, this means that the cumulative sum has exceeded the line with intercept C_A in Figure 8.1 and the site is judged contaminated. However, if the conclusion in step 4 is made, then the cumulative sum has fallen below the line with intercept C_B in Figure 8.1 and the site is found to be clean. Notice that the intercept values depend on the error rates (α and β), the proportion that is being tested (P_0), and the proportion where the false negative error rate is estimated (P_1). The slope of these lines is determined strictly by P_0 and P_1.

Box 8.3 presents an example application of sequential testing.

8.6 Grouping Samples in Sequential Analysis

Under the random sampling approach discussed in section 7.3, a large number of soil units are selected from the site at one time, and the laboratory analysis is conducted on each unit, one at a time. In many situations it will be more efficient for the laboratory to analyze the soil units in small batches or groups rather than one at a time. The sequential procedure can be modified easily to account for this type of laboratory analysis.

The quantities Cs, P_0, P_1, α, and β are defined in exactly the same way as for stratified sampling. Similarly, the stopping rules are also defined in exactly the same way. The only modification to the previously discussed procedures is in the calculation of k. Previously, after each soil unit was analyzed, k was calculated as the cumulative number of soil units that exceeded the cleanup standard, Cs. To modify k to take into account the grouped nature of the data, k should be computed as the cumulative number of soil units that exceed Cs after each batch has been analyzed. This minor modification is illustrated in Box 8.4 for groups of five, using the example of Box 8.3. In the example, after 4 groups of 5 or a total of 20 soil units, $k = 4 \geq nM + CA = 3.0324$, so sampling is terminated and the site is considered contaminated.

Box 8.3
An Example of Sequential Testing

Assume that for the chemical under investigation, the following values have been specified in the objectives worksheet: $\alpha = .05$, $\beta = .10$, $P_0 = .05$, and $P_1 = .02$. In this case, the quantities necessary to construct the acceptance and rejection regions are:

$$B = \ln(\frac{.05}{.9}) = -2.8904, \text{ and } A = \ln(\frac{.95}{.1}) = 2.2513;$$

$$R_1 = \frac{.95}{.98} = .9694, \text{ and } R_2 = \frac{.05}{.02} = 2.5; \; M = \frac{\ln(.9694)}{\ln(.9694/2.5)} = .0328;$$

$$C_B = \frac{-2.8904}{\ln(2.5/.9694)} = -3.0510, \text{ and } C_A = \frac{2.2513}{\ln(2.5/.9694)} = 2.3764;$$

$$n_f = \{ \frac{1.28\sqrt{.02} + 1.64\sqrt{.05}}{.05 - .02} \}^2 = 333$$

Below is a sequence of outcomes that might be observed for a particular chemical. Note that the values of the boundary limits use the values of M, CA, and CB, computed above. In the table, k = the cumulative number of soil units that are found to have excessive levels of the contaminant. The process terminates after the 18th soil unit has been analyzed. Prior to the 18th observation, the value of k falls between the computed values of nM+CA and nM+CB. However, with the 18th soil unit, k = 3 ≥ nM+CA = 2.9668, and hence the null hypothesis is accepted, i.e., the site fails to meet the cleanup standard.

Soil unit	Sample outcome	k	nM+CA	nM+CB	Decision
1	0	0	2.4092	-3.0182	continue
2	0	0	2.4420	-2.9854	continue
3	0	0	2.4748	-2.9526	continue
4	0	0	2.5076	-2.9198	continue
5	0	0	2.5404	-2.8870	continue
6	0	0	2.5732	-2.8542	continue
7	0	0	2.6060	-2.8214	continue
8	0	0	2.6388	-2.7886	continue
9	0	0	2.6716	-2.7558	continue
10	0	0	2.7044	-2.7230	continue
11	1	1	2.7372	-2.6902	continue
12	0	1	2.7700	-2.6574	continue
13	0	1	2.8028	-2.6246	continue
14	0	1	2.8356	-2.5918	continue
15	0	1	2.8684	-2.5590	continue
16	1	2	2.9012	-2.5262	continue
17	0	2	2.9340	-2.4934	continue
18	1	3	2.9668	-2.4606	accept

Box 8.4
Example of Sequential Test Using Grouped Samples

Example using the data of Box 8.3 after grouping soil units into groups of 5.

Soil unit	Group	Sample outcome	k	$nM + C_A$	$nM + C_B$	Decision
1	1	0		2.4092	-3.0182	
2	1	0		2.4420	-2.9854	
3	1	0		2.4748	-2.9526	
4	1	0		2.5076	-2.9198	
5	1	0	0	2.5404	-2.8870	continue
6	2	0		2.5732	-2.8542	
7	2	0		2.6060	-2.8214	
8	2	0		2.6388	-2.7886	
9	2	0		2.6716	-2.7558	
10	2	0	0	2.7044	-2.7230	continue
11	3	1		2.7372	-2.6902	
12	3	0		2.7700	-2.6574	
13	3	0		2.8028	-2.6246	
14	3	0		2.8356	-2.5918	
15	3	0	1	2.8684	-2.5590	continue
16	4	1		2.9012	-2.5262	
17	4	0		2.9340	-2.4934	
18	4	1		2.9668	-2.4606	
19	4	0		2.9996	-2.4278	
20	4	1	4	3.0324	-2.3950	accept

8.7 Summary

Sequential sampling means that a statistical test is performed after each sample or small batch of samples is collected and analyzed. Sequential testing does not imply that a time dynamic phenomenon is being monitored. Volume 2, which discusses ground water, considers sampling and analysis over time. Sequential sampling is performed during steady state conditions and is used only to reduce the sample size required for a decision.

Sequential sampling procedures terminate when enough evidence is obtained to either accept or reject the null hypothesis. Thus, sequential tests can respond quickly to very clean or very contaminated sites and in these cases require far less sampling than the conventional methods discussed in Chapter 7. In situations where contaminant concentrations at the site are only marginally different from the cleanup standard, the sequential procedure can be expected to require more samples until the sample size approaches the sample size required for the fixed sample procedure.

The procedure and some computational aspects of sequential testing are discussed. Sequential sampling and testing are treated separately from the discussions of other similar evaluation methods because of the distinct differences in sampling approach. However, the chapter makes comparisons with Chapter 7 procedures for sample size determination.

9. SEARCHING FOR HOT SPOTS

As suggested by Barth et al. (1989), it may be desirable to verify cleanups by documenting that no hot spots could be identified provided that a sampling plan was used that had an acceptably large probability of finding hot spots. This chapter discusses how to conduct a valid sampling program to search for hot spots and the conclusions that can be drawn regarding the presence or absence of hot spots. In general, the methods in this chapter are presented so they are easy to understand and apply.

This chapter first describes the literature that discusses methods for locating hot spots. This will provide the interested reader with an avenue into discussions regarding specific applications and details. A simple approach, useful under two different sampling designs, is summarized. This enables application of selected basic methods without having to obtain and study the literature.

9.1 Selected Literature that Describes Methods for Locating Hot Spots

Table 9.1 lists several references regarding hot spots and their identification. Gilbert (1987) offers a general overview of the hot spot searching technique, including example applications of the simplest methods as well as more advanced application. Zirschky and Gilbert (1984) offer applications of these methods at hazardous waste sites.

9.2 Sampling and Analysis Required to Search for Hot Spots

9.2.1 Basic Concepts

The term hot spot is used frequently in discussions regarding the sampling of hazardous waste sites, yet there is no universal definition of what constitutes a hot spot. The methods in this chapter model hot spots as localized elliptical areas with concentrations in excess of the cleanup standard. Hot spots are generally small relative to the area being sampled. The hot spot must either be considered a volume defined by the projection of the surface area through the soil zone that will be sampled or a discrete horizon within the soil zone that will be sampled. When a sample is taken and the concentration of a chemical exceeds the cleanup standard for that chemical, it is concluded that the sampling position in the field was located within a hot spot.

Table 9.1 Selected references regarding the methodologies for identifying hot spots at waste sites

Gilbert, R.O. (1982)	Some Statistical Aspects of Finding Hot Spots and Buried Radioactivity
Gilbert, R.O. (1987)	Statistical Methods for Environmental Pollution Monitoring
Parkhurst, D.F. (1984)	Optimal Sampling Geometry for Hazardous Waste Sites
Singer, D.A. (1972)	Elipgrid: A Fortran IV Program for Calculating the Probability of Success in Locating Elliptical Targets with Square, Rectangular, and Hexagonal Grids
Singer, D.A. (1975)	Relative Efficiencies of Square and Triangular Grids in the Search for Elliptically Shaped Resource Targets
USEPA (1985)	Verification of PCB Spill Cleanup by Sampling and Analysis
Zirschky, J. and Gilbert, R.O. (1984)	Detecting Hot Spots at Hazardous Waste Sites

Hot spot location techniques involve systematic sampling from a grid of sampling points arranged in a particular pattern. If a systematic sample is taken and none of the samples yield concentrations in excess of the cleanup standard, then no hot spots were found and the site is judged clean. However, what does this mean in terms of the chances of contaminant residuals remaining at the site? Since all of the soil could not be sampled, hot spots could still be present. An important question is: What level of certainty is there that no hot spots exist at the site? The answer to this question requires that several other questions be answered. For example:

- What shape hot spot is of concern: circular, fat-elliptical, skinny-elliptical?
- What is the length of the longest axis of the hot spot : 1 cm, 10 m, or 100 m?
- What sampling pattern was used: square, triangular?
- What was the distance between sampling points in the grid: 0.1 m, 1 m, 100 m?

If these questions are answered; a sampling plan implemented; and no hot spots are found, it is possible to conclude with an associated level of confidence that no hot spots of a certain size are present. In general, there is a smaller chance of detecting hot spots and less confidence in conclusions when:

- Hot spot sizes of interest become smaller;
- Hot spots are likely to be narrow;
- A square rather than a triangular grid is used; and
- The spacing between grid points is increased.

Figure 9.1 illustrates a sampling grid with hot spots of various sizes and shapes. Hot spots B and D were "hit" with sampling points and hot spots A anc C were missed.

If one of the samples results in concentrations in excess of the applicable cleanup standard, a hot spot has been identified. The conclusion is that the site is not clean. The normal, reasonable action will be to continue remediation in the areas identified as hot spots. However, once these locations are remediated, another systematic sample, over the entire site, with a new random start must be taken in order to conclude with confidence that no hot spots of a specified size and shape are present at the site. Because of this requirement it may be advisable, after identifying the presence of a single hot spot, to continue less formal searching followed by treatment throughout the entire sample area.

9.2.2 Choice of a Sampling Plan

The sampling plan requires no calculations. Instead all the information is obtained from tables. Figure 9.2 describes the grid spacing definition for two grid configurations and how to calculate the parameter for defining the ellipse shape.

The sampling plan for hot spot detection can be approached in three ways. The three factors listed in Table 9.2 control the performance of a hot spot detection sampling episode. Two of these factors are chosen and fixed. The third factor is determined by the choice of the first two factors. Table A.11 includes information that allows choice of two factors while providing the resulting third parameter.

Figure 9.1 A Square Grid of Systematically Located Grid Points with Circular and Elliptical Hot Spots Superimposed

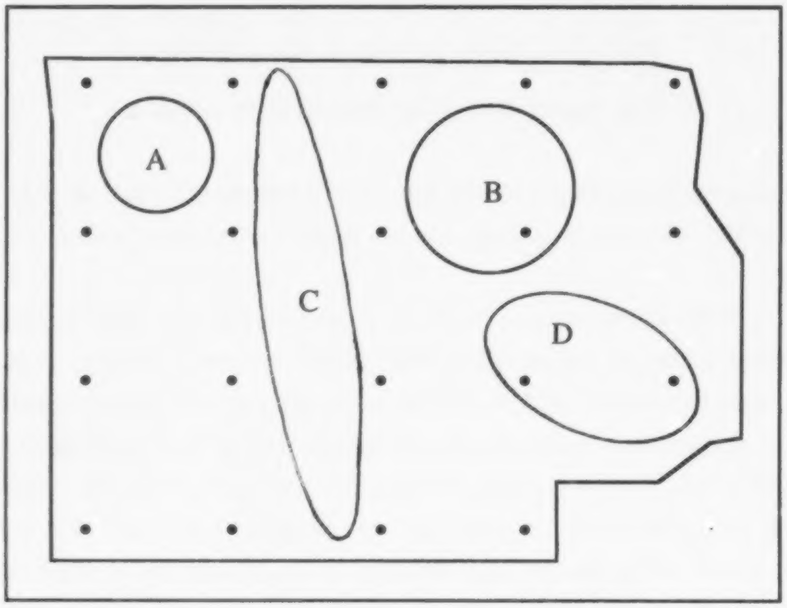


Figure 9.2 Grid Spacing and Ellipse Shape Definitions for the Hot Spot Search Table in Appendix A (Table A.11)

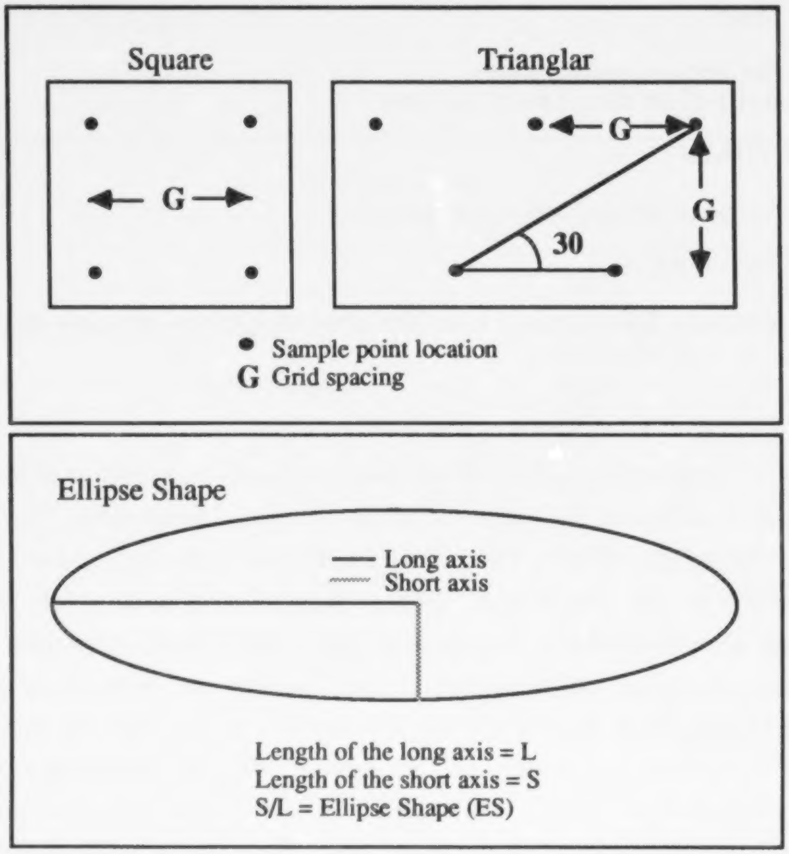

Square
Triangular
G
G
G
30
● Sample point location
G Grid spacing
Ellipse Shape
Long axis
Short axis
Length of the long axis = L
Length of the short axis = S
S/L = Ellipse Shape (ES)

Table 9.2 Factors controlling the design of a hot spot search sampling plan

GRID PATTERN	
	Spacing between sample points. Geometry of the sample point locations.
HOT SPOT SHAPE	
	The length of the long axis of the hot spot.
FALSE POSITIVE RATE	
	An acceptable false positive probability; concluding that no hot spots are present when there is at least one present.

Three examples are offered that describe the approaches to sample plan design. First, suppose that the size of the hot spot is known or assumed. The shape and size of the hot spots that are being searched for are elliptical with a long axis of $L = 5$ m and a short axis of $S = 2$ m. Therefore, the ellipse shape, $ES = S/L = 2/5 = 0.4$. In addition, the sampling team decided that they could accept no more than a 10 percent chance of missing a hot spot if a hot spot was present the false positive rate. A triangular grid pattern was chosen because the probability of detection was better with an elliptical shaped hot spot and the sampling team had experience laying out a triangular coordinate system. The triangular grid pattern in Table A.11 is entered for a value of $ES = 0.4$ across the top and a false positive rate of $\alpha = .10$ or less within the table. This corresponds to an L/G value of 0.9, since $L = 5$, and $0.9 = 5/G$, $G = 5.55$, or a grid spacing in a triangular pattern of 5.6 m. The density of the grid spacing must be evaluated with respect to the size of the sample area.

Once the grid spacing density has been determined it is important to estimate for the sample area how many samples would be required given sampling intervals of 5.6 m on a triangular grid as specified in Figure 9.2. The following method in Box 9.1 can be used to approximate the sample size necessary when area and grid interval are known.

Box 9.1
Approximating the Sample Size
When Area and Grid Interval Are Known

$$n = A/G^2$$

Where: n = total number of samples required

A = size of the area to be sampled (in the same units of measures as G)

G = grid spacing as defined in Figure 9.2

For example, suppose that a lagoon will be sampled that is 45 m by 73 m. This is a 3285-m^2 lagoon. The number of samples required is:

$$3285\ m^2 / (5.6\ m)^2 = 104$$

On the other hand, a lagoon that is 17 m by 20 m or 340-m^2 would require the following number of samples:

$$340\ m^2 / (5.6\ m)^2 = 11$$

If the size of the area is relatively small, then the level of confidence described above may be affordable and acceptable. However, if the area is large and the number of samples required excessive, alternatives are available.

For example, a second approach can be considered that limits the samples from the 3285-m^2 lagoon. Suppose that no more than 40 samples are available because of cost, time, or logistics. The minimum grid spacing is estimated to be:

$$3285\ m^2 / G^2 \leq 40$$

$$G \geq 9.1\ m$$

The question becomes: what probability statement can be made with a 9.1 m grid spacing searching for the same size hot spot. Review of Table A.11 indicates that if $L/G = 5/9.1 = .55$, and $ES = S/L = 2/5 = 0.4$ then $.33 < \alpha < .63$. Reference to Gilbert (1987) indicates that $\alpha \approx .55$. This means first that the cost has been reduced by taking 64 fewer samples from the 3285-sq. m lagoon. This was accomplished by increasing the grid spacing from 5.6 m to 9.1 m. However, the sampling cost reduction increases the chance of missing contamination. Specifically, the chance of missing a hot spot and concluding that the site is clean when a hot spot with an ES of 0.4 and a long axis of 5 m is really present increases from 10 percent to 55 percent when the sample size is reduced from 104 to 40. If this chance is unacceptably high, there is a third approach.

The third approach involves fixing the false positive rate, fixing the sample size or grid spacing, and searching for hot spots that are larger or have a different shape. Suppose it could be safely assumed that the hot spot of concern was not as elliptically shaped or as skinny as the ellipse with an $ES = 0.4$. Instead, the $ES = L = 4/5 = 0.8$. The long axis remained at 5 m, but the short axis doubled from 2 m to 4 m. For the grid spacing of $G = 5.6$ m, the $L/G = 5/5.6 = 0.9$. From Table A.11 it is clear that the false positive rate is low, $\alpha = .01$. A willingness to search for a larger sized or fatter shaped hot spot improves the performance of the hot spot search technique from a 10 percent false positive rate to a less than 1 percent false positive rate with no increase in sample intensity above 104 samples.

9.2.3 Analysis Plan

The analysis is straightforward. Establish a grid of sampling points as described in Chapter 5 with density and pattern determined using the methods in section 9.2.2 and Figure 9.2. If one of the chemical measurement results exceeds the cleanup standard then conclude that a hot spot has been found and the completion of remediation can not be verified. If none of the samples exceeds the cleanup standard, assume that the site is clean and conclude with the level of confidence associated with the sampling plan that it is unlikely a hot spot exists at the site.

9.3 Summary

Hot spots are generally defined as relatively small, localized, elliptical areas with contaminant concentrations in excess of the cleanup standard. Samples that are taken and found to exceed the cleanup standard are defined as being located within a hot spot.

Locating hot spots involves systematic sampling from a grid of sampling points arranged in a specific pattern. Several questions must be answered to conclude with a level of confidence that no hot spots of a certain size are present:

- What size hot spot is of concern?
- What sampling pattern was used?
- What was the distance between sampling points in the grid?

The sampling plan for hot spot detection is guided by the dimensions and shape of the grid pattern, the hot spot shape of interest, and the false positive rate. The information needed is contained in Table A.11. Three illustrative examples present sampling plans for these cases:

- The size of the hot spot and false positive rate are known or assumed, and the grid spacing/sample size is determined;
- Sample size/grid spacing and ellipse shape are fixed, and the false positive rate is determined;
- The false positive rate and sample size or grid spacing are fixed, and hot spot size is determined.

10. THE USE OF GEOSTATISTICAL TECHNIQUES FOR EVALUATING THE ATTAINMENT OF CLEANUP STANDARDS

The science of geostatistics involves the analysis of spatially correlated data. There are several features of geostatistics that are important to any potential user.

- Geostatistical methods provide a powerful and attractive method for mapping spatial data. Geostatistical methods provide for interpolation between existing data points that have been collected in a spatial array and a method for estimating the precision of the interpolation.
- Geostatistical methods are complicated mathematically, and the procedures required to contour an area cannot be practically implemented by hand and calculator.
- New users of geostatistics will need to devote time to understanding the basic approach, concepts and the unique vocabulary associated with geostatistical methods.
- To help explore applications, PC-based geostatistical computer software is now readily available to the EPA community (USEPA, 1988). However, some preliminary study should be completed, and then the software can be used as an educational and exploratory tool to better understand how geostatistical methods perform.

This chapter:

- Explains fundamental concepts regarding geostatistical methods;
- Offers a point of departure into the literature that will provide more details;
- Discusses which cleanup scenarios can benefit the most from a geostatistical evaluation;
- Describes which geostatistical methods are most appropriate for evaluating the completion of cleanup; and
- Lists software available for implementing geostatistical methods.

CHAPTER 10: THE USE OF GEOSTATISTICAL TECHNIQUES FOR EVALUATING THE ATTAINMENT OF CLEANUP STANDARDS

10.1 Background

10.1.1 What Is Geostatistics and How Does It Operate?

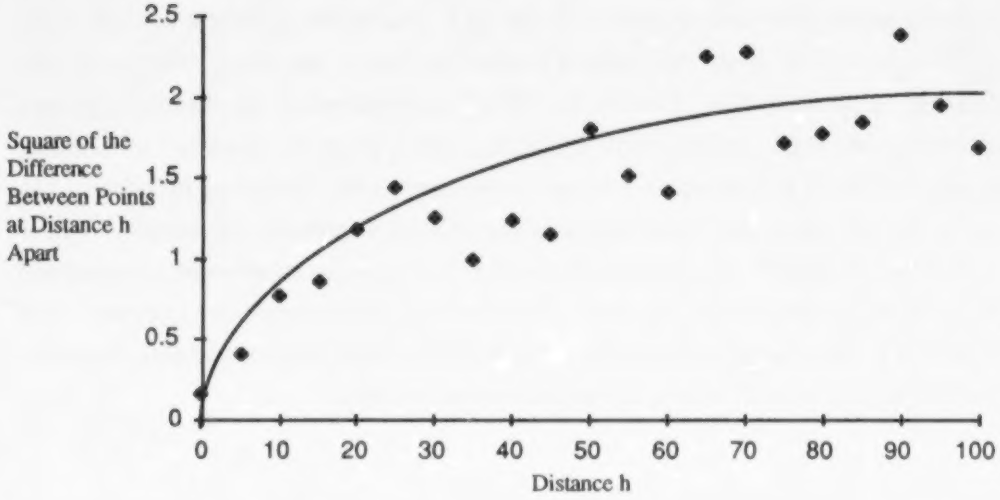
Many view the science of geostatistics in a broad context as the use of statistical methods applied to the geographic and geological sciences. Others refer to geostatistics as a science that strictly applies to the family of methods that enable the analysis, evaluation, or characterization of spatially correlated data. Regardless, kriging and variogram modeling are primary tools of geostatistical analysis.

In simple terms, a geostatistical analysis can be viewed as a two-step process. First, a model is developed that predicts the spatial relationship between a location where a concentration will be estimated and the existing data obtained from sample points which are various distances away from the location. Existing data points nearer to the location will tend to be closely related and have a large influence on the estimate, and points far away will tend to be less related and, therefore, impose less influence. This relationship function, which describes how influential nearby existing data will be, is modeled and called a variogram or semi-variogram.

Figure 10.1 illustrates the general form of a standard or typical variogram model. The X or horizontal axis measures the distance between sample points. The vertical or Y axis measures the degree of relationship between points. When there is little distance between points it is expected that there will be little variability between points. As the distance between points increases, the difference or variability between points increases. The form of this relationship depends on what the variogram modeler knows about characteristics of the site and the data, and what assumptions are reasonable to make regarding spatial relationships at the site.

The second step of the geostatistical analysis is kriging. This involves estimating chemical concentrations for each point or block in the area of concern. For each point to be estimated, the surrounding points provide a weighted contribution to the estimate. The weightings are determined by using the variogram model, the location of the point that is being estimated, and the proximity of other nearby data values, enabling chemical concentration estimation for locations within the sample area that were not sampled and therein lies the true value of a geostatistical analysis. In addition to estimates of the concentration, kriging allows estimation of the precision associated with the estimate. If the surrounding data are highly variable, or if the closest data points are relatively far away, the precision may be low.

Figure 10.1 An Example of an Empirical Variogram and a Spherical Variogram Model



Kriging provides concentration and associated precision estimates across the site at all possible points or blocks within the site. The concentration and precision estimates can then be graphically contoured across the site. Maps, plotting concentration isopleths, are the final product. In addition, a precision map that provides isopleths of the kriging variance or some function of the kriging variance is generated. These sorts of maps are illustrated in Flatman and Yfantis (1984) and USEPA (1987b).

As a slightly more technical conclusion to this section, consider the following discussion of kriging and variogram modeling. Kriging is an interpolation method based on a weighted moving average where the weights are assigned to samples in a way that minimizes the variance associated with interpolated estimates. The estimation variance is computed as a function of the spatial relationship model known as the variogram, the location of the sampling points relative to each other, and to the location being estimated (USEPA, 1988).

10.1.2 Introductory Geostatistical References

The discussion in section 10.1.1 is intended to provide a simple notion of how kriging operates. The next level of understanding requires that the reader consult specialized literature and a practicing geostatistician. Several general discussions of geostatistics are available and are listed in Table 10.1. In addition to the references in Table 10.1, there is a wide range of refereed journal literature supporting the theory and application of geostatistics. Finally, the EPA's Environmental Monitoring Systems Laboratory in Las Vegas, Nevada (EMSL-LV), includes a group of researchers specializing in the application of geostatistical methods to environmental monitoring problems. The group is responsible for the development of the GEOEAS software referenced in Tables 10.1, 10.3, and Box 10.1. In addition, the EMSL-LV has produced refereed literature and funded university researchers. The researchers operating under cooperative agreement with the EMSL-LV have produced a series of reports that also provide insights regarding application of geostatistical methods to environmental problems.

10.2 Soils Remediation Technology and the Use of Geostatistical Methods

As recognized in Chapter 1, there are a variety of soils remediation methods. Geostatistical methods have many applications, and are especially useful during remedial investigations where a primary objective is to characterize the extent of contamination. Geostatistical techniques, particularly specialized kriging techniques referenced in section 10.3, will also be useful for evaluating certain soils remediation efforts.

This section provides guidance that will help in deciding whether geostatistical methods are most appropriate for use under different types of soils remediation methods. The reader should note that in cases where geostatistical approaches are not necessarily called for if they are used then the geostatistical approaches will give the same result as the classical approaches used throughout the document. The choice of whether to use geostatistical data analysis and evaluation methods depends on the physical arrangement of the cleanup system, its mode of operation, and the effect that the remediation technology will have on the soils environment.

Table 10.1 Selected introductory and advanced references that introduce and discuss geostatistical concepts

INTRODUCTORY	
Clark, I. (1979)	Practical Geostatistics
Davis, J.C. (1986)	Statistical and Data Analysis in Geology
USEPA (1987a)	Data Quality Objectives for Remedial Response Activities: Development Process
USEPA (1987b)	Data Quality Objectives for Remedial Response Activities: Example Scenario RI/FS Activities at a Site with Contaminated Soils and Ground Water
USEPA (1988)	GEOEAS (Geostatistical Environmental Assessment Software) User's Guide
ADVANCED	
Journel, A.G. and Huijbregts, C.H. (1978)	Mining Geostatistics
David, M. (1984)	Geostatistical Ore Reserve Estimation
Verly, G. et al. (1984)	Geostatistics for Natural Resources Characterization

10.2.1 Removal

Soils remediation may involve either permanent or temporary removal of soils. Soils may be permanently transported away from the site. However, soils may be temporarily removed to undergo treatment and then returned. In these situations, geostatistical methods may be useful for efficiently directing the removal.

For example, although a single three-dimensional geostatistical study or a series of two-dimensional geostatistical studies at various depth horizons would have been preferred during the site characterization phase, this may not have been done. Therefore, during removal, as the surface material is skimmed off and new layers are exposed, the areas of greatest concentration may change. This changing condition with depth could be characterized via a geostatistical study. However, there are practical requirements in this situation. In order to be most successful and efficient onsite rapid chemical analysis and geostatistical data analysis must take place.

A geostatistical analysis will permit the estimation of concentrations between the sampled points and allow prediction of which areas should and should not be removed. As horizons are reached that are below the cleanup standard, they can be avoided. The sampling program and data analysis have the ability to operate in a useful and constructive way that will help direct the cleanup effort and minimize costs. Indicator and probability kriging, discussed below, are ideal candidates for evaluating areas that are above and below cleanup standards.

10.2.2 Treatment Involving Homogenization

Many soils remediation technologies homogenize the soils media. This occurs during soils fixation or chemical modification when soil mixers are used to blend materials with the soil media. Sampling this type of process could occur at a discharge point of the mixing apparatus. In this instance, samples may be taken, placed in canisters, and allowed to solidify or undergo the chemical reaction. After an established period of time, the media in the canisters can be extracted and the leachate concentrations tested relative to the appropriate cleanup standard. Samples may also be acquired onsite after the mixing equipment such as banks of steam injection augers, has passed over each location that has been pre-selected for sampling to test attainment of the cleanup standard.

Regardless of how the sampling is conducted, from a statistical perspective, there are several anticipated results. First, there should be reduction in the variability of chemical contaminants across the site. One way of viewing the effect of treatment is that it has reduced the magnitude of the large values in the distribution of values at the site. This can be thought of as "bringing in" the upper tail of the distribution such that the distribution becomes less lognormal-like and more bell-shaped or normal-like. In practical terms, this is the same as reducing the variance. In short, the site should be more homogenous, and there should be a more random behavior of contaminants across the site. Finally, the degree of spatial relationship will be reduced because of the homogeneity. That is, a point 1m away from a point of concern should be just as similar as a point 50m away.

Because of these anticipated results, geostatistical applications are less useful when remediation results in a homogenization. First, it is likely that the spatial correlation has been grossly disturbed by the treatment process. Also, sampling may occur at a discharge point or in association with the operation of a mixing device, rather than in a spatial framework. If the treatment technology is operating as anticipated, the effectiveness will be high; the extractable concentrations will be low relative to the cleanup standard and will have a small variance. Under this scenario, a sampling and analysis program as discussed in Chapters 4-9 can be implemented with a minimum of samples to verify the effectiveness of treatment rather than require an elaborate geostatistical study.

10.2.3 Flushing

There is a family of soils remediation techniques that can be thought of as flushing methods. They rely on surface manifolds attached to extraction wells on one end and to suction pumps on the other end. These systems can be designed to remove infiltrated water, artificial liquids, or air. In either case, the liquid or air is the media used to transport the contaminants. The liquid can flush out soluble contaminants, and the air can flush out volatile contaminants. Often extraction systems have to contend with both air and liquid.

A system of extraction wells, screened at appropriate depths, are installed across the contaminated area. Each of the wells is linked by a manifold or piping system, which is connected to a pump system that provides the vacuum for withdrawal. The dynamics of removal differ depending on many factors including the makeup of the soils media, the degree of infiltration, the surrounding ground water system, the type of contaminants, and the media that is being extracted. Regardless of these factors, there is a tendency with these systems to create zones of influence around each well. Depending on how long the system has operated and many other factors, the zone of influence will have much higher or lower concentrations than the surrounding area. The site will then tend to have a series of zones of influence across the site. Some of the zones will overlap; others will be irregular in shape because of irregularities in the soils media or the turning on and off of banks of wells in the system.

Geostatistical methods are generally not practical for characterizing sites that have been remediated using flushing technologies because of the highly complex structure associated with the many overlapping zones of influence around each of the extraction wells that are distributed across the site. Although it may be technically possible to geostatistically model this structure, many samples would be required to provide sufficient resolution of the many complex gradients across the site.

However, it may be that by the time verification sampling is conducted the zones of influence are not likely to be apparent and the site is anticipated to be uniformly below the relevant cleanup standard. If extraction has been completed to this point and there is interest in characterizing the concentration profile across the site, a geostatistical study may be warranted. However, the main objective at this stage will normally not be to characterize the extent of the remaining contaminants that have concentrations below the cleanup standard, but instead to simply document that the site has met its cleanup objectives.

10.3 Geostatistical Methods that Are Most Useful for Verifying the Completion of Cleanup

As previously described, there are many methods of variogram modeling and many approaches to kriging. Each technique requires different assumptions or has advantages in a particular application. The traditional forms of kriging allow estimates of central tendency and variance throughout an area. These forms, which include simple, ordinary, and universal kriging, require different assumptions regarding the model used to make the kriging estimates. These types of kriging methods can be used to describe the extent of contamination remaining and the precision associated with the concentration estimates. In this way, the traditional forms of kriging are useful for cleanup verification.

In addition to the more common methods of kriging described above, there are several forms of nonparametric kriging, such as indicator and probability kriging, that have been developed relatively recently and are directly useful for evaluating attainment of cleanup standards. These types of kriging are the best forms of kriging for demonstrating that a particular area is less than a cleanup standard, and unlike the conventional forms of kriging, these forms are distribution-free.

Indicator kriging operates basically by kriging data that have been transformed into zeros and ones. For each measurement, the value is transformed to a zero if the measurement was less than or equal to the cleanup standard, and transformed to a one if the measurement was greater than the cleanup standard. The data set of zeros and ones is then used to produce kriging estimates of the probability of exceeding the cleanup standard across the site. It then becomes possible to produce a map that contours the probabilities of having concentrations in excess of the cleanup standard. Extensions of indicator kriging to probability kriging allow the development of false positive and false negative error maps. That is, probability kriging can be used to estimate where there is a chance that an area that appears to be clean is actually dirty and where there is a chance that areas that might be indicated dirty are actually clean. Figures 10.2, 10.3, and 10.4 were adapted from the probability kriging study of a lead smelter (Flatman et al., 1985).

Although these forms of kriging are directly applicable to the cleanup verification problem, they are relatively new methods. Nonparametric and Baysian kriging are currently an active area of research. Understanding and application of these kriging methods will require a substantial investment of time and study. Table 10.2 offers some initial references.

10.4 Implementation of Geostatistical Methods

As mentioned in the introduction to this chapter, kriging cannot be conveniently or practically implemented without a computer and the appropriate software. Even with the appropriate software, it will take an interested individual a considerable investment of time to understand the jargon and mathematics associated with geostatistical methods.

Table 10.2 Introductory references for indicator, probability, and nonparametric global estimation kriging

Buxton, B.E. (1985)	Geostatistical Construction of Confidence Intervals for Global Reserve Estimation
Isaaks, E.H. (1984)	Risk Qualified Mappings for Hazardous Waste Sites: A Case Study in Distribution Free Geostatistics
Journel, A.G. (1983)	Nonparametric Estimation of Spatial Distributions
Sullivan, J. (1984)	Conditional Recovery Estimation Through Probability Kriging

In many cases, it is best to recognize the power and utility of a geostatistical study and acquire, or at least have available, the expertise of a geostatistician. An alternative is to obtain a first-level familiarity with the methodology and then use a softwarepackage along with example data sets to explore the practical dynamics and effects of different modeling decisions.

The EMSL-LV has recently produced the first version of a geostatistical software package that provides a convenient environment for exploring the application of geostatistical methods to hazardous waste site sampling problems (USEPA, 1988). The software operates on a PC and is provided in an executable form. It is entirely in the public domain and can be obtained using the information in Box 10.1.

The software does not support indicator and probability kriging at this point; however, as the software undergoes development, it is anticipated that these will be added.

There are other geostatistical software packages available in the public domain that can be purchased. Table 10.3 lists some examples and sources of software.

Figure 10.2 Contour Map of the Probability in Percent of Finding the Value of 1,000 ppm or a Larger Value

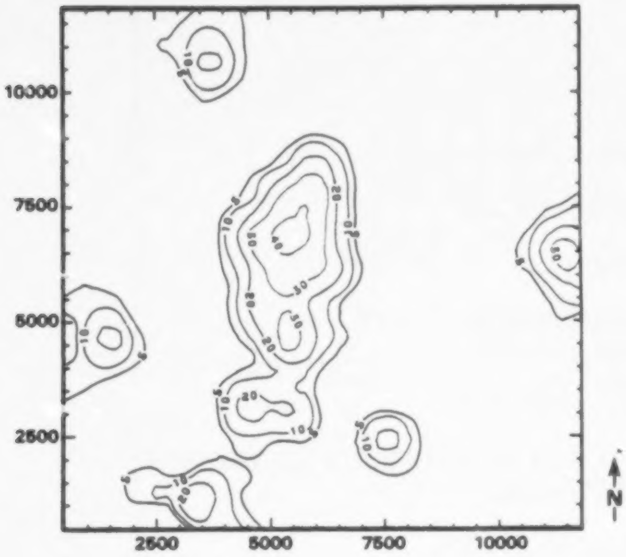

10000
7500
5000
2500
2500
5000
7500
10000
N

Figure 10.3 Contour Map of the Probability in Percent of a False Positive in the Remedial Action Areas and the 1,000 ppm Contour Line

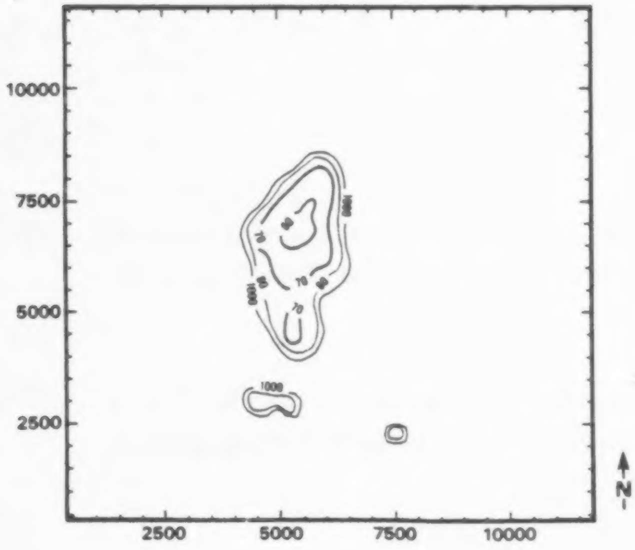
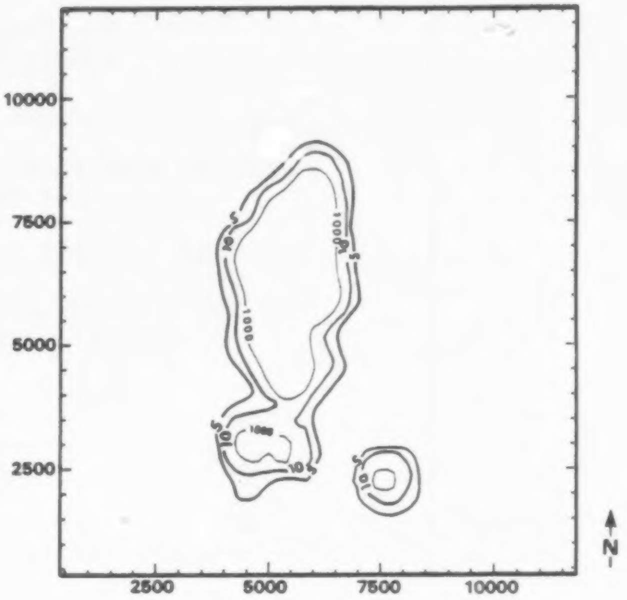

10000
7500
5000
2500
2500
5000
7500
10000
N

Figure 10.4 Contour Map of the Probability in Percent of a False Negative in the Remedial Action Areas and the 1,000 ppm Contour Line


10000
7500
5000
2500
2500
5000
7500
10000
N

Box 10.1

Steps for Obtaining Geostatistical Software from EMSL-LV

The software:

- Operates on a PC;
- Is provided in an executable form;
- Is entirely in the public domain; and
- Can be obtained by writing to:

Evan Englund (GEO-EAS)
USEPA, EMSL-LV, EAD
P.O. Box 93478
Las Vegas, NV 89193-3478

PLEASE, YOU MUST DO THE FOLLOWING TO OBTAIN THE SOFTWARE!:

1) PRE-FORMAT ALL DISKETTES.

2) SEND ENOUGH DISKETTES FOR 3 MEGABYTES OF STORAGE AS FOLLOWS:

TYPE		NUMBER
5 1/4"	1.2MB	3
5 1/4"	360KB	9
3 1/2"	1.44MB	3
3 1/2"	722KB	6

10.5 Summary

Geostatistical methods provide a method for mapping spatial data that enables both interpolation between existing data points and a method for estimating the precision of the interpolation.

Geostatistical applications normally involve a two-step process. First, a spatial correlation model is developed that predicts how much spatial relationship exists among sample points various distances apart.

Table 10.3 Selected geostatistical software

Program	Source
Geo-EAS SYSTEM (Geostatistical Environmental Assessment Software)	See Box 10.1
USGS Statpac Programs	COGS (Computer Oriented Geological Society) P.O. Box 1317 Denver, Colorado 80201-1317
TOXIPAC	Geostat Systems International, Inc. P.O. Box 1193 Golden, CO 80402
GEOBASE and GEORES	GEOMATH 4860 Ward Road Wheat Ridge, CO 80033

The second step, kriging, involves estimating chemical concentrations for locations within the sample area that were not sampled. For each point to be estimated, the surrounding points provide a weighted contribution to the estimate based on the variogram model, the location of the point being estimated, and the proximity of other nearby data values. Kriging also allows estimation of the precision associated with the estimated chemical concentrations. Maps that plot concentration isopleths are the final product of the geostatistical analysis.

Geostatistical methods have many applications in soil remediation technology, especially when the extent of contamination needs to be characterized. This chapter includes guidance to help decide whether geostatistical data analysis and evaluation methods are appropriate for use with three types of soils remediation activities: removal, treatment involving renamed homogenization, and flushing.

Of the many methods of variogram modeling and many approaches to kriging, each requires different assumptions or has advantages in certain applications. The traditional forms of kriging, including simple, ordinary, and universal, are primarily useful for characterization but may also be used for cleanup verification. Nonparametric, indicator, and probability kriging are the best forms for demonstrating probabilistically that an area is less than a cleanup standard and, unlike the traditional forms, are distribution-free.

Geostatistical techniques referred to in the chapter will need in-depth study by the intended user before being applied. References are provided to help familiarize the interested reader. Because kriging cannot be conveniently or practically implemented without a computer and the appropriate software, a first-level familiarity with the methodology along with use of a software package is a practical way of exploring example applications and data sets. EPA has developed the first version of a geostatistical software for the novice, available by following instructions at the end of this chapter.

BIBLIOGRAPHY

Armitage, P. 1947. "Some Sequential Tests of Student's Hypothesis." *Journal of the Royal Statistical Society*. Series B, Vol. 9 : 250-263.

Barich, J. Memo to B. Johnson on April 21, 1988.

Barnett, V., and T. Lewis. 1978. "Outliers in Statistical Data." John Wiley & Sons, Bath, England.

Barth, D. S., Mason, B. J., Starks, T. H., and K. W. Brown. 1989. Soil Sampling Quality Assurance User's Guide, Second Edition, USEPA Cooperative Agreement No. CR 814701, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

Bavinck, H. F., Roels, J. M., and J. J. Vegter. 1988. The Importance of Measurement Procedures in Curative and Preventive Soil Protection." In K. Wolf, W. J. van den Brink, L. F. J. Colon (eds.) Contaminated Soil '88, 125-133, Kluwer Academic Publishers.

Beckett, P. H. T., and R. Webster. 1979. "Soil Variability: A Review." *Soils and Fertilizers*. Vol. 34, No. 1: 1-15.

Benson, R. C., Glaccum, R. A., and M. R. Noel. 1984. Geophysical Techniques for Sensing Buried Wastes and Waste Migration. Miami, Fla.: Technos, [June 1984].

Bisgaard, S., and W. G. Hunter. 1986. Report No. 7, Studies in Quality Improvement: Designing Environmental Regulations, Center for Quality and Productivity Improvement, University of Wisconsin-Madison, [February 1986].

Bishop, T. A. 1985. Statistical View of Detection and Determination Limits in Chemical Analyses Prepared for the Committee on Applications of Statistical Techniques in Chemical Problems. Columbus, Ohio: Battelle Columbus Laboratories, [January 5, 1982].

Box, G. E. P., Hunter W. G., and J. S. Hunter. 1987. Statistics For Experimenters. New York: John Wiley.

Bross, I. D. 1985. "Why Proof of Safety is Much More Difficult Than Proof of Hazard." *Biometrics*. Vol. 41: 785-793.

Brown, K. W., and S. C. Black. 1983. "Quality Assurance and Quality Control Data Validation Procedures Used for the Love Canal and Dallas Lead Soil Monitoring Programs." *Environmental Monitoring and Assessment*. Vol. 3: 113-122.

BIBLIOGRAPHY

Brown, G. H., and N. I. Fisher. 1972. "Subsampling a Mixture of Sampled Material." *Technometrics*, Vol. 14, No. 3: 663-668.

Brown, M. B., and R. A. Wolfe. 1983. "Estimation of the Variance of Percentile Estimates." *Computational Statistics and Data Analysis.* Vol. 1: 167-174.

Brownlee, K. A. 1965. Statistical Theory and Methodology in Science and Engineering, 2nd. New York: John Wiley & Sons, Inc.

Burgess, T. M., and R. Webster. 1980a. "Optimal Interpolation and Isarithmic Mapping of Soil Properties: I. The Semi-Variogram and Punctual Kriging." *Journal of Soil Science.* Vol. 31: 315-331.

Burgess, T. M., and R. Webster. 1980b. "Optimal Interpolation and Isarithmic Mapping of Soil Properties: II. Block Kriging." *Journal of Soil Science.* Vol. 31: 333-341.

Buxton, B. E. 1985. Geostatistical Construction of Confidence Intevals for Global Reserve Estimation, Unpublished Master's Thesis, Stanford University.

Campbell, J. B. 1979. "Spatial Variability of Soils." *Annals of the Association of American Geographers.* Vol. 69, No. 4: 544-556.

Casey, D., Nemetz, P. N., and D. Uyeno. 1985. "Efficient Search Procedures for Extreme Pollutant Values." *Environmental Monitoring and Assessment.* Vol. 5: 165-176.

Clark, I. 1979. Practical Geostatistics. New York: Elsevier Science.

Clayton, C. A., Hines, J. W., and P. D. Elkins. 1987. Detection Limits with Specified Assurance Probabilities. *Analytical Chemistry.* Vol. 59, No. 20: 2506-2514.

Cline, M. G. 1944. "Principles of Soil Sampling." *Soil Science.* Vol. 58: 275-288.

Cochran, W. 1977. Sampling Techniques. New York: John Wiley.

Cohen, A. C. 1961. "Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples." *Technometrics.* Vol. 3, No. 4: 535-541.

Conover, W. J. 1980. Practical Nonparametric Statistics. New York: John Wiley.

D'Agostino, R. B. 1970. "A Simple Portable Test of Normality: Geary's Test Revisited." *Psychological Bulletin.* Vol.. 74, No. 2: 138-140.

David, M. 1984. Geostatistical Ore Reserve Estimation.

Davis, J. C. 1986. Statistical Data Analysis in Geology. John Wiley and Sons, New York.

BIBLIOGRAPHY

Doctor, P. G., and R. O. Gilbert. 1978. "Two Studies in Variability for Soil Concentrations: With Aliquot Size and with Distance." Selected Environmental Plutonium Reports of the NAEG. Richland, Wash.: Battelle Pacific Northwest Laboratory, [June 1978].

Dowd, P. A. 1982. "Lognormal Kriging - The General Case." *Mathematical Geology*. Vol. 14, No. 5: 475-499.

DuMouchel, W. H., Govindarajulu, Z., and E. Rothman. 1973. "Note on Estimating the Variance of the Sample Mean in Stratified Sampling." *The Canadian Journal of Statistics*. Vol. 1, No. 2: 267-274.

Duncan, A. J. 1962. "Bulk Sampling: Problems and Lines of Attack." *Technometrics*. Vol. 4, No. 2: 319-344.

Elder, R. S., Thompson, W. O., and R. H. Myers. 1980. "Properties of Composite Sampling Procedures." *Technometrics*. Vol. 22, No. 2: 179-186.

Esposito, M. P., McArdle, J. L., Crone, A. H., Greber, J. S., Clark, R., Brown, S., Hallowell, J. B., Langham, A., and C. D. McCandlish. 1985. Guide for Decontaminating Buildings, Structures, and Equipment at Superfund Sites. Columbus, Ohio: Battelle Columbus Laboratories, [March 1985].

Exner, J. H., Keffer, W. D., Gilbert, R. O., and R. R. Kinnison. 1985. "Sampling Strategy for Remedial Action at Hazardous Waste Sites: Cleanup of Soil Contaminated by Tetrachlorodibenzo-P-Dioxin." *Hazardous Waste and Hazardous Materials*. Vol. 2, No. 4: 503-521.

Filliben, J. J. 1975. "Probability Plot Correlation Coefficient Test for Normality." *Technometrics*. Vol. 17, No. 1: 111-117.

Finney, D. J. 1948. "Random and Systematic Sampling in Timber Surveys." *Forestry*. Vol. 22: 64-99.

Flatman, G. T., Brown, K. T., and J. W. Mullins. 1985. "Probabilistic Spatial Contouring of the Plume Around a Lead Smelter." The 6th National Conference on Management of Uncontrolled Hazardous Waste Sites, Superfund '85 Proceedings, Hazardous Material Control Research Institute: 442-448.

Flatman, G. T., and A. A. Yfantis. 1984. "Geostatistical Strategy for Soil Sampling: The Survey and the Census." *Environmental Monitoring and Assessment*. Vol. 4: 335-349.

Ford, P., and P. Turina. 1985. Characterization of Hazardous Waste Sites--A Methods Manual, Volume I--Site Investigations. Las Vegas, Nevada: EPA Environmental Monitoring Systems Laboratory, [April 1985].

Garner, F. C. 1985. Comprehensive Scheme for Auditing Contract Laboratory Data [interim report]. Las Vegas, Nevada: Lockheed-EMSCO.

Gastwirth, J. L., and H. Rubin. 1971. "Effect of Dependence on the Level of Some One-Sample Tests." *Journal of the American Statistical Association*. Vol. 66: 816-820.

Gilbert, R. O. 1982. *Tran-Stat: Statistics for Environmental Studies*, No. 19, Some Statistical Aspects of Finding Hot Spots and Buried Radioactivity. Richland, Wash.: Pacific Northwest Laboratory.

Gilbert, R. O. 1983. *Tran-Stat: Statistics for Environmental Studies*, No. 24, Field Sampling Designs, Simple Random and Stratified Random Sampling. Richland, Wash.: Battelle Pacific Northwest Labs.

Gilbert, R. O. 1984. *Tran-Stat: Statistics for Environmental Studies*, No. 26, Field Sampling Designs: Systematic Sampling. Richland, Wa.: Battelle Memorial Institute, Pacific Northwest Laboratory.

Gilbert, R. O. 1987. Statistical Methods for Environmental Pollution Monitoring. New York: Van Nostrand Reinhold.

Gilbert, R. O., and P. G. Doctor. 1985. "Determining the Number and Size of Soil Aliquots for Assessing Particulate Contaminant Concentrations." *Journal of Environmental Quality*. Vol. 14: 286-292.

Gilbert, R. O., and R. R. Kinnison. 1981. "Statistical Methods for Estimating the Mean and Variance from Radionuclide Data Sets Containing Negative, Unreported or Less-Than Values." *Health Physics*. Vol. 40: 377-390.

Gilbert, R. O., Miller, M. L, and H. R. Meyer. 1989. "On the Design of a Sampling Plan to Verify Compliance with EPA Standards for Radium-226 in Soil at Uranium Mill Tailings Remedial-Action Sites." ASA/EPA Conference Proceedings on Interpretation of Environmental Data. IV Compliance Sampling, October 5-6, 1987 (EPA-230-03-047).

Gilliom, R. J., and D. R. Helsel. 1986. "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data, 1. Estimation Techniques." *Water Resources Research*. Vol. 22, No. 2: 135-146.

Gleit, A. 1985. "Estimation for Small Normal Data Sets with Detection Limits." *Environmental Science and Technology*. Vol. 19, No. 12: 1201-1206.

Grubbs, F. E. 1969. "Procedures for Detecting Outlying Observations in Samples." *Technometrics*. Vol. 11, No. 1: 1-21.

Guttman, I. 1970. Statistical Tolerance Regions: Classical and Bayesian. No. 26 Griffin's Statistical Monographs and Courses. Alan Stuart, ed, Darien, Conn.: Hafner Publishing.

Hammond, L. C., Pritchett, W. L., and V. Chew. 1958. "Soil Sampling in Relation to Soil Heterogeneity." *Soil Science Society of America Proceedings*. Vol. 22: 548-552.

Hansen, M., Hurwitz, W, and W. Madow. 1953. Sample Survey Methods and Theory, Volume 1. New York: John Wiley.

Hazardous Materials Control Research Institute. 1985. 6th National Conference on Management of Uncontrolled Hazardous Waste Sites. Washington, D.C.: HMCRI, [November 4-6, 1985].

Hazardous Materials Control Research Institute. 1986. 7th National Conference on Management of Uncontrolled Hazardous Waste Sites. Washington, D.C.: HMCRI, [December 1-3, 1986].

Helsel, D. R., and R. J. Gilliom. 1986. "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data, 2. Verification and Applications." *Water Resources Research.* Vol. 22, No. 2: 147-155.

Hoaglin, D. C., Mosteller, F., and J. W. Tukey. 1983. Understanding Robust and Exploratory Data Analysis. New York: John Wiley.

Hughes, J. P., and D. P. Lettenmaier. 1981. "Data Requirements for Kriging: Estimation and Network Design." *Water Resources Research.* Vol. 17, No. 6: 1641-1650.

Isaaks, E. H. 1984. Risk Qualified Mappings for Hazardous Waste Sites: A Case Study in Distribution Free Geostatistics. Master's Thesis. Stanford University.

Journel, A. G. 1983. "Nonparametric estimation of spatial distributions." *Mathematical Geology.* Vol. 15, No. 3: 445-468.

Journel, A. G. 1984. Indicator Approach to Toxic Chemical Sites [EPA Project No. CR-811235-02-0-1984]. Stanford, Ca.: Stanford University.

Journel, A. G., and C. J. Huijbregts. 1978. Mining Geostatistics. New York: Academic Press, London.

Kish, L. 1965. Survey Sampling. New York: John Wiley.

Koop, J. C. 1976. Systematic Sampling of Two-Dimensional Surfaces and Related Problems. Research Triangle Park, N.C.: Research Triangle Institute, [May 1976].

Krige, D. G., and E. J. Magri. 1982. "Geostatistical Case Studies of the Advantages of Lognormal-De Wijsian Kriging With Mean for a Base Metal Mine and a Gold Mine." *Mathematical Geology.* Vol. 14, No. 6:547-555.

Land, C. E. 1971. "Confidence Intervals for Linear Functions of the Normal Mean and Variance." *Annals of Mathematical Statistics.* Vol. 42, No.4: 1187-1205.

Land, C. E. 1975. Tables of Confidence for Linear Functions of the Normal Mean and Variance. Selected Tables in Mathematical Statistics, Vol. III. pp.385-419. Providence, R.I.: American Mathematical Society.

Legg, D. E., Shufran, K. A., and K. V. Yeargan. 1985. "Evaluation of Two Sampling Methods for Their Influence on the Population Statistics of Alfalfa Weevil (Coleoptera: Curculionidae) Larva Infestations in Alfalfa." *Journal of Economic Entomology*. Vol. 78, No. 6: 1468-1474.

Lehmann, E. L. 1975. Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Madow, W. G., and L. H. Madow. 1944. "On the Theory of Systematic Sampling." *Annals of Mathematical Statistics*. Vol. 15: 1-24.

Mage, D. T. 1982. "Objective Graphical Method for Testing Normal Distributional Assumptions Using Probability Plots." *American Statistician*. Vol. 36, No. 2: 116-120.

Mason, B. J. 1983. Preparation of Soil Sampling Protocol: Techniques and Strategies. Las Vegas, Nevada: Environmental Research Center, UNLV, [August 1983].

McBratney, A. B., and R. Webster. 1981. "Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables - II. Program and Examples." *Computers and Geosciences*. Vol. 7, No. 4: 335-365.

McBratney, A. B. Webster, R., and T. M. Burgess. 1981. "Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables - I. Theory and Method." *Computers and Geosciences*. Vol. 7, No. 4: 331-334.

Mee, R. W. 1984. "Tolerance Limits and Bounds for Proportions Based on Data Subject to Measurement Error." *Journal of Quality Technology*. Vol. 16, No. 2: 74-80.

Mee, R. W. 1989. "Normal Distribution Tolerance Limits for Stratified Random Samples." *Technometrics*. Vol. 31, No. 1: 99-105.

Mee, R. W., Owen, D. B., and J. C. Shyu. 1986. "Confidence Bounds for Misclassification Probabilities Based on Data Subject to Measurement Error." *Journal of Quality Technology*. Vol. 18, No. 1: 29-40.

Mendenhall, W., and L. Ott. 1980. Understanding Statistics. N. Scituate, Mass.: Duxbury Press.

Milne, A. 1959. "Centric Systematic Area-Sample Treated as a Random Sample." *Biometrics*. Vol. 15, No. 2: 270-297.

Natrella, M. 1963. Experimental Statistics. Washington, D.C.: U.S. Department of Commerce, National Bureau of Standards.

Neter, J., Wasserman, W., and G. Whitmore. 1982. Applied Statistics. Boston: Allyn-Bacon.

BIBLIOGRAPHY

Osborne, J. G. 1942. "Sampling Errors of Systematic and Random Surveys of Cover-Type Areas." *Journal of the American Statistical Association*. Vol. 37: 256-264.

Ower, D. B. 1963. Factors for One-Sided Tolerance Limits and for Variables Sampling Plans. Albuquerque, N.M.: Sandia Corporation.

Palley, M. N., and L. G. Horwitz. 1961. "Properties of Some Random and Systematic Point Sampling Estimators." *Forest Science*. Vol. 77: 52-65.

Parkhurst, D. F. 1984. "Optimal Sampling Geometry for Hazardous Waste Sites." *Environmental Science and Technology*. Vol. 18, No. 7: 521-523.

Patel, J. K. 1986. "Tolerance Limits- A Review." *Communications Statistics: Theory and Methods*. Vol. 15, No. 9: 2719-2762.

Peshkova, N. V. 1970. "Comparison of Results of Determinations of the Density of Plant Populations by Systematic and by Restricted Random Sampling Procedures." *Ecology (Ekologiya)*. Vol. 2: 180-181.

Phillip, G. M, and D. F. Watson. 1986. "Matheronian Geostatistics - Quo Vadis?" *Mathematical Geology*. Vol. 18, No. 1: 93-117.

Puente, C. E., and R. L. Bras. 1986. "Disjunctive Kriging, Universal Kriging, or No Kriging: Small Sample Results with Simulated Fields." *Mathematical Geology*, Vol. 18, No. 3: 287-305.

Rao, P. V., Rao, P. S. C., Davidson, J. M., and L. C. Hammond. 1979. "Use of Goodness-of-Fit Tests for Characterizing the Spatial Variability of Soil Properties." *Soil Scientist Society of America Journal*. Vol. 43, No. 2: 274-278.

Rao, R. G. S., and F. T. Ulaby. 1977. "Optimal Spatial Sampling Techniques for Ground Truth Data in Microwave Remote Sensing of Soil Moisture." *Remote Sensing of Environment*. Vol. 6, No. 4: 289-301.

Rendu, J. M. 1979. "Normal and Lognormal Estimation." *Mathematical Geology*. Vol. 11, No. 4: 407-422.

Ripley, B. D. 1981. Spatial Statistics. New York: John Wiley.

Roels, J. M., and P. J. Jonker. 1983. "Probability Sampling Techniques for Estimating Soil Erosion." *Soil Science Society of America Journal*. Vol. 47: 1224-1228.

Rohde, C. A. 1976. "Composite Sampling." *Biometrics*. Vol. 32: 273-282.

Rohde, C. A. 1979. Batch, Bulk, and Composite Sampling. Sampling Biological Populations, pp. 365-367. Edited by R.M. Cormack. Fairland, Md.: International Cooperative Pub. House.

Schaeffer, D. J., Derster, H. W., and D. G. Janardan. 1980. Grab versus composite sampling: A primer for managers and engineers. *J. Environ. Mgmt.* Vol. 4: 157-163.

Schaeffer, D.J., and D. G. Janardan. 1978. Theoretical comparison of grab and composite sampling programs. *Biom J.* Vol. 20: 215-227.

Scheaffer, R. L., Mendenhall, W., and L. Ott. 1979. Elementary Survey Sampling, Second Edition. Boston: Duxbury Press.

Schmid, C. F. 1983. Statistical Graphics: Design Principles and Practices, John Wiley and Sons, Inc.

Schwartz, J. E. 1985. "Neglected Problem of Measurement Error in Categorical Data." *Sociological Methods and Research.* Vol. 13, No. 4: 435-466.

Schweitzer, G. E., and J. A. Santolucito, editors. 1984. Environmental Sampling for Hazardous Wastes. ACS Symposium Series 267. Washington, D.C.: American Chemical Society.

Shapiro, S. S., and M. B. Wilk. 1965. "Analysis of Variance Test for Normality (Complete Samples)." *Biometrika.* Vol. 52: 591-611.

Sharpe, K. 1970. "Robustness of Normal Tolerance Intervals." *Biometrika.* Vol. 57, No. 1: 71-78.

Sigmund, D. 1985. Sequential Analysis: Tests and Confidence Intervals. New York: Springer-Verlag.

Singer, D. A. 1972. "Elipgrid. A Fortran IV Program for Calculating the Probability of Success in Locating Elliptical Targets With Square, Rectangular, and Hexagonal Grids." *Geological Programs.* Vol. 4: 1-16.

Singer, D. A. 1975. Relative Efficiencies of Square and Triangular Grids in the Search for Elliptically Shaped Resource Targets. *Jour Research U. S. Geological Survey.* Vol. 3, No. 2: 163-167.

Sirjaev, A. N. 1973. Statistical Sequential Analysis. Providence, R.I.: American Mathematical Society.

Size, W. B., editor. 1987. Use and Abuse of Statistical Methods in the Earth Sciences. New York: Oxford University Press.

Snedecor, G. W., and W. G. Cochran. 1980. Statistical Methods. Seventh Edition Ames Iowa: The Iowa State Press.

Sokal, R. R., and F. J. Rohlf. 1973. Introduction to Biostatistics. San Francisco: W. H. Freeman.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry: The Principles and Practice of Statistics in Biological Research. Second Edition. New York: W. H. Freeman.

BIBLIOGRAPHY

Starks, T. H. 1986. "Determination of Support in Soil Sampling." *Mathematical Geology*. Vol. 18, No.6: 529-537.

Stellingwerf, D. A., and S. Lwin. 1985. "Stratified Sampling Compared with Two-Phase Stratified Cluster Sampling for Timber Volume Estimation." *Netherlands Journal of Agricultural Science*. Vol. 33, No. 2: 151-160.

Sullivan, J. 1984. Conditional Recovery Estimation Through Probability Kriging. Doctoral Thesis, Stanford University.

Swinger, D. A. 1972. Elipgrid, A Fortran IV Program for Calculating the Probability of Success in Locating Elliptical Targets with Square, Rectangular, and Hexogonal Grids. *Geostatistical Programs*. Vol. 4: 1-16.

Switzer, P. 1983. When Will a Pollutant Standard Be Exceeded: Model Prediction and Uncertainty. Technical Report No. 67. New Canaan, Ct.: SIMS, [January 1983].

Taylor, J. A., Friend, G. R., and M. L. Dudzinski. 1984. "Influence of Sampling Strategy on the Relationships Between Fauna and Vegetation Structure, Plant Lifeform and Floristics." *Australian Journal of Ecology*. Vol. 9: 281-287.

Thomas, V. W., and R. R. Kinnison. 1983. Recommended Sampling Strategies for Spatial Evaluation of Windblown Contamination Around Uranium Tailings Piles. Richland, Wa.: Battelle Pacific Northwest Labs, [October 1983].

Tornqvist, L. 1963. "Theory of Replicated Systematic Cluster Sampling With Random Start." *Review of the International Statistical Institute*. Vol. 31, No. 1: 11-23.

Tukey, J. W. 1977. Exploratory Data Analysis. Reading, Mass.: Addison-Wesley.

U.S. Environmental Protection Agency. 1983. Preparation of Soil Sampling Protocol: Techniques and Strategies, Washington D.C., August 1983 (EPA 600/4-83-020).

U.S. Environmental Protection Agency. 1984. Soil Sampling Quality Assurance User's Guide, First Edition, Washington, D.C., May 1984 (EPA 600/4-84-043).

U.S. Environmental Protection Agency. 1985. Verification of PCB Spill Cleanup by Sampling and Analysis, Washington D.C., August 1985 (EPA 560/5-85-026).

U.S. Environmental Protection Agency. 1986a. A Primer on Kriging. Washington, D.C.

U.S. Environmental Protection Agency. 1986b. Guidance Document for Cleanup of Surface Impoundment Sites, Washington D.C., June 1986 (OSWER Directive 9380.0-6).

U.S. Environmental Protection Agency. 1986c. Guidance on Remedial Actions for Contaminated Ground Water at Superfund Sites [draft] Washington D.C., October 1986.

U.S. Environmental Protection Agency. 1986d. Superfund Public Health Evaluation Manual, Washington D.C.: EPA [October 1986].

U.S. Environmental Protection Agency. 1987a. Data Quality Objectives For Remedial Response Activities: Development Process, Washington D.C., March 1987 (EPA 540/G-87/003).

U.S. Environmental Protection Agency. 1987b. Data Quality Objectives For Remedial Response Activities: Example Senario RI/FS Activities at a Site with Contaminated Soils and Ground Water, Washington D.C., March 1987 (EPA 540/G-87/004).

U.S. Environmental Protection Agency. 1987c. EPA Journal, The New Superfund: Protecting People and the Environment, Washington D.C., Vol. 13, No. 1.

U.S. Environmental Protection Agency. 1987d. Surface Impoundment Clean Closure Guidance Manual [draft], Washington D.C., March 1987, (OSWER Directive 9476.0-8.C).

U.S. Environmental Protection Agency. 1987e. Test Methods for Evaluating Solid Waste, Washington D.C., November 1987 (SW-846).

U.S. Environmental Protection Agency. 1988. GEOEAS (Geostatistical Environmental Assessment Software User's Guide, EMSL-Las Vegas, NV, (EPA 600/4-88/033).

Verhagen, E. J. H. 1988. "Review of Thermal and Extraction Soil Treatment Plants in the Netherlands," National Institute of Public Health and Environmental Protection (RIVM) Laboratory for Waste and Emission Research (LAE), In K. Wolf, W. J. vanden Brink, F. J. Colon (eds.), Contaminated Soil '88, 797-808, Kluwer Academic Publishers.

Verly, G., David, M., Journel, A. G., and A. Marachal, editors. 1984. Geostatistics for Natural Resources Characterization, Parts 1 and 2, D. Reidel Publishing Company.

Wald, A. 1973. Sequential Analysis. New York: Dover Publications.

Watson, G. S. 1984. "Smoothing and Interpolation by Kriging and with Splines." *Mathematical Geology*. Vol. 16, No. 6: 601-615.

Webster, R., and T. M. Burgess. 1980. "Optimal Interpolation and Isarithmic Mapping of Soil Properties: III. Changing Drift and Universal Kriging." *Journal of Soil Science*. Vol. 31: 505-524.

Wesolowsky, G. O. 1976. Multiple Regression and Analysis of Variance. New York: John Wiley.

Wetherill, G. B. 1975. Sequential Methods in Statistics. New York: Halsted Press.

Willits, N. H. 1982. Compliance Testing for Statistical Air Quality Standards. Technical Report No. 64. New Canaan, Ct.: SIMS, [October 1982].

Wolter, K. M. 1984. "Investigation of Some Estimators of Variance for Systematic Sampling." *Journal of the American Statistical Association*. Vol. 79, No. 388: 781-790.

Yakowitz, S. J., and F. Szidarovsky. 1985. "A Comparison of Kriging with Nonparametric Regression Methods." *Journal of Multivariate Analysis*. Vol. 16: 21-53.

Yates, F. 1981. Sampling Methods for Censuses and Surveys, Fourth Edition. New York: Oxford University Press.

Yfantis, E. A., Flatman, G. T., and J. V. Behar. 1987. "Efficiency of Kriging Estimation for Square, Triangular, and Hexogonal Grids." *Journal of Mathematical Geology*. Vol. 19, No. 3: 183-205.

Zirschky, J., and R. O. Gilbert. 1984. "Detecting Hot Spots at Hazardous Waste Sites." *Chemical Engineering*. Vol. 91: 97-100.

APPENDIX A: STATISTICAL TABLES

Table A.1 Table of t for selected alpha and degrees of freedom

Use alpha to determine which column to use based on the desired parameter, $t_{1-\alpha,df}$ or $t_{1-\alpha/2,df}$. Use the degrees of freedom to determine which row to use. The t value will be found at the intersection of the row and column. For values of degrees of freedom not in the table, interpolate between those values provided.

		α for determining $t_{1-\alpha,df}$							
		.25	.10	.05	.025	.01	.005	.0025	.001
		α for determining $t_{1-\alpha/2,df}$							
		.50	.20	.10	.05	.02	.01	.005	.002
	df								
	1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309
	2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327
Degrees of Freedom df	3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215
	4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173
	5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893
	6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208
	7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785
	8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501
	9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297
	10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144
	11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025
	12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930
	13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852
	14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787
	15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733
	16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686
	17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646
	18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610
	19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579
	20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552
	21	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527
	22	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505
	23	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485
	24	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467
	25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450
	26	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435
	27	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421
	28	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408
	29	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396
	30	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385
	40	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307
	60	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232
	120	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160
	400	0.675	1.284	1.649	1.966	2.336	2.588	2.823	3.111
	infinite	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090

APPENDIX A: STATISTICAL TABLES

Table A.2 Table of z for selected alpha or beta

Use alpha or beta to determine which row to read. Obtain the z value from the $z_{1-\alpha}$ or $z_{1-\beta}$ column adjacent to the desired α or β value.

β α	$z_{1-\beta}$ $z_{1-\alpha}$
0.450	0.124
0.400	0.253
0.350	0.385
0.300	0.524
0.250	0.674
0.200	0.842
0.100	1.282
0.050	1.645
0.025	1.960
0.010	2.326
0.0050	2.576
0.0025	2.807
0.0010	3.090

APPENDIX A: STATISTICAL TABLES

Table A.3 Table of k for selected alpha, P_0, and sample size where alpha = 0.10 (i.e., 10%)

Use alpha to determine which table to read. The k for use in a tolerance interval test is at the intersection of the column with the specified P_0 and the row with the sample size, n.

n	P_0			
	0.25	0.1	0.05	0.010
2	5.842	10.253	13.090	18.500
3	2.603	4.258	5.311	7.340
4	1.972	3.188	3.957	5.438
5	1.698	2.742	3.400	4.666
6	1.540	2.494	3.092	4.243
7	1.435	2.333	2.894	3.972
8	1.360	2.219	2.754	3.783
9	1.302	2.133	2.650	3.641
10	1.257	2.066	2.568	3.532
11	1.219	2.011	2.503	3.443
12	1.188	1.966	2.448	3.371
13	1.162	1.928	2.402	3.309
14	1.139	1.895	2.363	3.257
15	1.119	1.867	2.329	3.212
16	1.101	1.842	2.299	3.172
17	1.085	1.819	2.272	3.137
18	1.071	1.800	2.249	3.105
19	1.058	1.782	2.227	3.077
20	1.046	1.765	2.208	3.052
21	1.035	1.750	2.190	3.028
22	1.025	1.737	2.174	3.007
23	1.016	1.724	2.159	2.987
24	1.007	1.712	2.145	2.969
25	1.000	1.702	2.132	2.952
26	0.992	1.691	2.120	2.937
27	0.985	1.682	2.109	2.922
28	0.979	1.673	2.099	2.909
29	0.973	1.665	2.089	2.896
30	0.967	1.657	2.080	2.884
35	0.942	1.624	2.041	2.833
40	0.923	1.598	2.010	2.793
50	0.894	1.559	1.965	2.735
70	0.857	1.511	1.909	2.662
100	0.825	1.470	1.861	2.601
200	0.779	1.411	1.793	2.514
500	0.740	1.362	1.736	2.442
infinity	0.674	1.282	1.645	2.326

Table A.4 Table of k for selected alpha, P_0, and sample size where alpha = 0.05 (i.e., 5%)

n	P0			
	0.25	0.1	0.05	0.010
2	11.763	20.581	26.260	37.094
3	3.806	6.155	7.656	10.553
4	2.618	4.162	5.144	7.042
5	2.150	3.407	4.203	5.741
6	1.895	3.006	3.708	5.062
7	1.732	2.755	3.399	4.642
8	1.618	2.582	3.187	4.354
9	1.532	2.454	3.031	4.143
10	1.465	2.355	2.911	3.981
11	1.411	2.275	2.815	3.852
12	1.366	2.210	2.736	3.747
13	1.328	2.155	2.671	3.659
14	1.296	2.109	2.614	3.585
15	1.268	2.068	2.566	3.520
16	1.243	2.033	2.524	3.464
17	1.220	2.002	2.486	3.414
18	1.201	1.974	2.453	3.370
19	1.183	1.949	2.423	3.331
20	1.166	1.926	2.396	3.295
21	1.152	1.905	2.371	3.263
22	1.138	1.886	2.349	3.233
23	1.125	1.869	2.328	3.206
24	1.114	1.853	2.309	3.181
25	1.103	1.838	2.292	3.158
26	1.093	1.824	2.275	3.136
27	1.083	1.811	2.260	3.116
28	1.075	1.799	2.246	3.098
29	1.066	1.788	2.232	3.080
30	1.058	1.777	2.220	3.064
35	1.025	1.732	2.167	2.995
40	0.999	1.697	2.125	2.941
50	0.960	1.646	2.065	2.862
70	0.911	1.581	1.990	2.765
100	0.870	1.527	1.927	2.684
200	0.809	1.450	1.837	2.570
500	0.758	1.385	1.763	2.475
infinity	0.674	1.282	1.645	2.326

Table A.5 Table of k for selected alpha, P_0, and sample size where alpha = 0.01 (i.e., 1%)

n	P0			
	0.25	0.1	0.05	0.010
2	58.939	103.029	131.426	185.617
3	8.728	13.995	17.370	23.896
4	4.715	7.380	9.083	12.387
5	3.454	5.362	6.578	8.939
6	2.848	4.411	5.406	7.335
7	2.491	3.859	4.728	6.412
8	2.253	3.497	4.258	5.812
9	2.083	3.240	3.972	5.389
10	1.954	3.048	3.738	5.074
11	1.853	2.898	3.556	4.829
12	1.771	2.777	3.410	4.633
13	1.703	2.677	3.290	4.472
14	1.645	2.593	3.189	4.337
15	1.595	2.521	3.102	4.222
16	1.552	2.459	3.028	4.123
17	1.514	2.405	2.963	4.037
18	1.481	2.357	2.905	3.960
19	1.450	2.314	2.854	3.892
20	1.423	2.276	2.808	3.832
21	1.399	2.241	2.766	3.777
22	1.376	2.209	2.729	3.727
23	1.355	2.180	2.694	3.681
24	1.336	2.154	2.662	3.640
25	1.319	2.129	2.633	3.601
26	1.303	2.105	2.606	3.566
27	1.287	2.085	2.581	3.533
28	1.273	2.065	2.558	3.502
29	1.260	2.047	2.536	3.473
30	1.247	2.030	2.515	3.447
35	1.195	1.957	2.430	3.334
40	1.154	1.902	2.364	3.249
50	1.094	1.821	2.269	3.125
70	1.020	1.722	2.153	2.974
100	0.957	1.639	2.056	2.850
200	0.868	1.524	1.923	2.679
500	0.794	1.430	1.814	2.540
infinity	0.674	1.282	1.645	2.326

Table A.6 Sample sizes required for detecting a scaled difference tau of the mean from the cleanup standard for selected values of alpha and beta*

τ	$\beta = 0.20$			$\beta = 0.10$		
	α			α		
	0.10	0.05	0.01	0.10	0.05	0.01
0.05	1,798	2,470	4,020	2,621	3,422	5,213
0.10	449	618	1,005	655	856	1,303
0.15	200	274	447	291	380	579
0.20	112	154	251	164	214	326
0.25	72	99	161	105	137	209
0.30	50	69	112	73	95	145
0.35	37	50	82	53	70	106
0.40	28	39	63	41	53	81
0.45	22	30	50	32	42	64
0.50	18	25	40	26	34	52
0.55	15	20	33	22	28	43
0.60	12	17	28	18	24	36
0.65	11	15	24	16	20	31
0.70	9	13	21	13	17	27
0.75	8	11	18	12	15	23
0.80	7	10	16	10	13	20
0.85	6	9	14	9	12	18
0.90	6	8	12	8	11	16
0.95	5	7	11	7	9	14
1.00	4	6	10	7	9	13

*See section 6.1 and Box 6.3 for definitions of alpha (α), beta (β), and tau (τ).

Table A.7 Sample size required for test for proportions with $\alpha = .01$ and $\beta = .20$, for selected values of P_0 and P_1

Value of P under the alternative hypothesis, P_1

P_0	0.002	0.005	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
0.005	4,519										
0.010	1,131	3,383									
0.020	407	659	1,676								
0.030	241	333	577	2,649							
0.040	169	217	323	823	3,593						
0.050	129	158	218	434	1,058	4,515					
0.060	103	124	162	281	538	1,287	5,416				
0.070	86	101	127	202	340	639	1,509	6,295			
0.080	73	85	104	156	240	396	737	1,726	7,155		
0.090	64	73	88	125	182	276	451	833	1,938	7,994	
0.100	56	64	75	104	144	207	311	504	925	2,145	8,813

Value of P under the alternative hypothesis, P_1

P_0	0.010	0.020	0.050	0.100	0.150	0.200	0.250	0.300	0.350	0.400	0.450
0.050	218	434									
0.100	75	104	311								
0.150	43	53	103	469							
0.200	28	34	55	140	606						
0.250	21	24	35	71	171	723					
0.300	16	18	25	43	83	197	819				
0.350	12	14	19	30	50	93	217	894			
0.400	10	11	14	22	33	54	101	233	950		
0.450	8	9	11	16	24	36	58	106	243	986	
0.500	6	7	9	13	17	25	37	60	109	248	1,001

Table A.8 Sample size required for test for proportions with $\alpha = .05$ and $\beta = .20$, for selected values of P_0 and P_1

Value of P under the alternative hypothesis, P_1

P_0	0.002	0.005	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
0.005	2,623										
0.010	633	1,990									
0.020	222	373	986								
0.030	129	185	332	1,588							
0.040	90	119	183	485	2,171						
0.050	68	86	122	252	630	2,741					
0.060	55	67	90	162	317	772	3,297				
0.070	45	54	70	116	198	380	910	3,840			
0.080	38	45	57	88	139	234	441	1,044	4,371		
0.090	33	39	48	71	105	162	268	500	1,175	4,889	
0.100	29	34	41	58	83	120	183	301	557	1,303	5,394

Value of P under the alternative hypothesis, P_1

P_0	0.010	0.020	0.050	0.100	0.150	0.200	0.250	0.300	0.350	0.400	0.450
0.050	122	252									
0.100	41	58	183								
0.150	23	29	59	282							
0.200	15	19	31	83	368						
0.250	11	13	20	41	103	440					
0.300	8	10	14	25	49	119	500				
0.350	7	7	10	17	29	56	132	548			
0.400	5	6	8	12	20	33	61	142	583		
0.450	4	5	6	9	14	21	35	64	149	606	
0.500	3	4	5	7	10	15	23	37	67	153	616

Table A.9 Sample size required for test for proportions with $\alpha = .10$ and $\beta = .20$, for selected values of P_0 and P_1

Value of P under the alternative hypothesis, P_1

P_0	0.002	0.005	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
0.005	1,822										
0.010	426	1,398									
0.020	145	254	693								
0.030	84	124	229	1,133							
0.040	58	79	125	341	1,559						
0.050	44	57	82	175	447	1,975					
0.060	35	44	60	111	223	551	2,381				
0.070	29	35	47	79	138	269	652	2,778			
0.080	24	29	38	60	97	164	314	750	3,166		
0.090	21	25	32	48	72	113	189	357	846	3,544	
0.100	19	22	27	39	57	84	129	214	399	940	3,913

Value of P under the alternative hypothesis, P_1

P_0	0.010	0.020	0.050	0.100	0.150	0.200	0.250	0.300	0.350	0.400	0.450
0.050	82	175									
0.100	27	39	129								
0.150	15	20	41	202							
0.200	10	12	22	59	265						
0.250	7	9	14	29	73	318					
0.300	5	6	10	18	35	85	363				
0.350	4	5	7	12	21	40	95	398			
0.400	3	4	5	9	14	23	44	103	424		
0.450	3	3	4	6	10	15	25	47	108	441	
0.500	2	2	3	5	7	11	16	26	48	111	449

Table A.10 Tables for determining critical values for the exact binomial test, with α = 0.01, 0.05, and 0.10

To determine the critical value, select the column for P_0 specified in the attainment objectives, reading down the column finding the first number greater than the sample size, n, move up one row, read $r_{\alpha:n}$, the critical value, in the leftmost column.

Alpha = .01

P_0, Proportion of contaminated soil units under the null hypothesis

$r_{\alpha:n}$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12
0	459	228	152	113	90	75	64	56	49	44	40	37
1	662	330	219	164	130	108	92	81	71	64	58	53
2	838	418	277	207	165	137	117	102	91	81	74	67
3	1001	499	332	248	198	164	140	122	109	97	88	81
4	1157	577	383	287	229	190	162	142	126	113	102	93
5	1307	652	433	324	259	215	184	160	142	127	116	106
6	1453	725	482	360	288	239	204	178	158	142	129	118
7	1596	796	529	396	316	263	225	196	174	156	141	129
8	1736	866	576	431	344	286	244	213	189	170	154	141
9	1874	935	622	465	371	309	264	230	204	183	166	152
10	2010	1003	667	499	398	331	283	247	219	197	178	163

Alpha = .05

P_0, Proportion of contaminated soil units under the null hypothesis

$r_{\alpha:n}$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12
0	299	149	99	74	59	49	42	36	32	29	26	24
1	473	236	157	117	93	78	66	58	51	46	42	38
2	628	313	208	156	124	103	88	77	68	61	56	51
3	773	386	257	192	153	127	109	95	84	76	69	63
4	913	456	303	227	181	150	129	112	100	89	81	74
5	1049	523	348	261	208	173	148	129	115	103	93	85
6	1182	590	392	294	234	195	167	146	129	116	105	96
7	1312	655	436	326	260	217	185	162	143	129	117	107
8	1441	719	478	358	286	238	203	178	158	142	128	117
9	1568	782	521	390	311	259	221	193	172	154	140	128
10	1693	845	562	421	336	280	239	209	185	167	151	138

Alpha = .10

P_0, Proportion of contaminated soil units under the null hypothesis

$r_{\alpha:n}$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12
0	230	114	76	57	45	38	32	28	25	22	20	19
1	388	194	129	96	77	64	55	48	42	38	34	31
2	531	265	176	132	105	88	75	65	58	52	47	43
3	667	333	221	166	132	110	94	82	73	65	59	54
4	798	398	265	198	158	132	113	98	87	78	71	65
5	926	462	308	230	184	153	131	114	101	91	83	76
6	1051	525	349	262	209	174	149	130	115	104	94	86
7	1175	587	390	292	234	194	166	145	129	116	105	96
8	1297	648	431	323	258	215	184	160	142	128	116	106
9	1418	708	471	353	282	235	201	175	156	140	127	116
10	1538	768	511	383	306	255	218	190	169	152	138	126

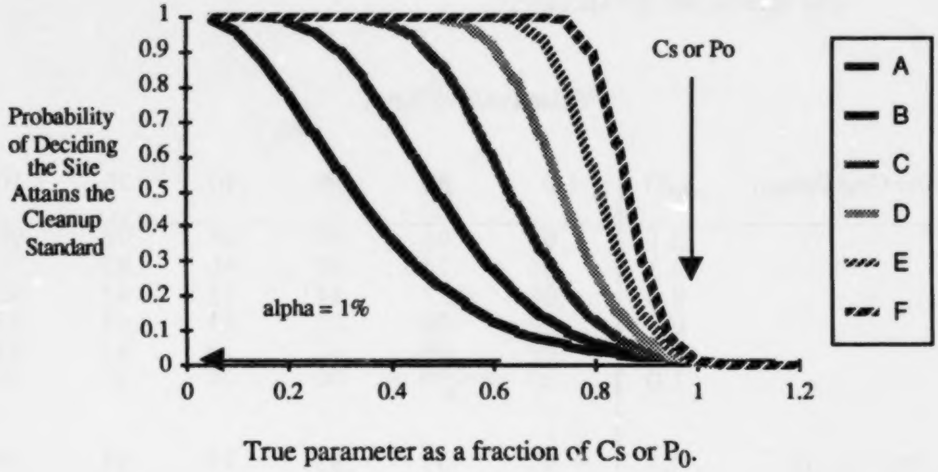
Table A.11 The false positive rates associated with hot spot searches as a function of grid spacing and hot spot shape

False Positive Rates

		ES					
Triangular Grid Pattern	L/G	1.0	.80	.60	.40	.20	.10
	0.1	.95	.96	.97	.98	.98	.99
	0.3	.66	.74	.80	.86	.93	.96
	0.5	.08	.27	.44	.63	.82	.91
	0.7	.00	.00	.08	.33	.65	.83
	0.9	.00	.00	.00	.10	.47	.72
	1.0	.00	.00	.00	.04	.37	.66
Square Grid Pattern	0.1	.97	.97	.98	.98	.98	.99
	0.3	.72	.77	.80	.88	.94	.97
	0.5	.21	.38	.54	.69	.85	.92
	0.7	.00	.02	.16	.42	.70	.85
	0.9	.00	.00	.00	.17	.53	.76
	1.0	.00	.00	.00	.08	.44	.70

Source: These tables were extracted from the graphs in Gilbert (1987).

Figure A.1 Power Curves for $\alpha = 1\%$



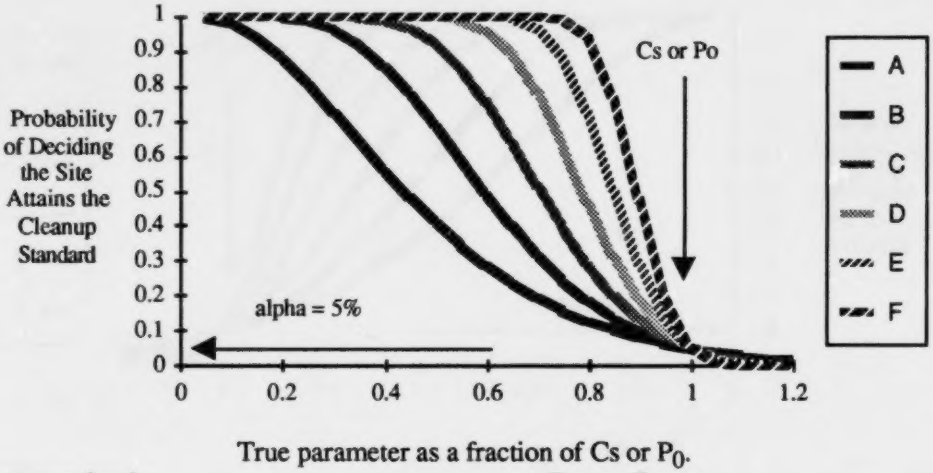
Parameters for the Power Curves	Power Curve: A	B	C	D	E	F
$\alpha =$	.01	.01	.01	.01	.01	.01
$\beta =$	.20	.20	.20	.20	.20	.20
$\mu_1 =$	.19*Cs	.36*Cs	.53*Cs	.65*Cs	.75*Cs	.81*Cs
$P_1 =$	$.19{*}P_0$	$.36{*}P_0$	$.53{*}P_0$	$.65{*}P_0$	$.75{*}P_0$	$.81{*}P_0$

Approximate sample sizes for simple random sampling for testing the parameters indicated

Parameters being tested	Power Curve: A	B	C	D	E	F
Mean						
with cv(data) = .5	4	7	12	21	41	70
with cv(data) = 1	16	25	46	82	161	279
with cv(data) = 1.5	35	56	103	185	362	626
Proportions						
$P_0 = 10\%$						
Non-parametric test	101	179	356	670	1353	2384
Tolerance Intervals	16	38	89	184	399	728
$P_0 = 20\%$						
Non-parametric test	46	81	161	301	607	1066
Tolerance Intervals	12	26	60	122	261	473

Note: α = saying the site is clean when dirty, β = saying the site is dirty when clean, $1-\beta$ = saying the site is clean when clean.

Figure A.2 Power Curves for $\alpha = 5\%$



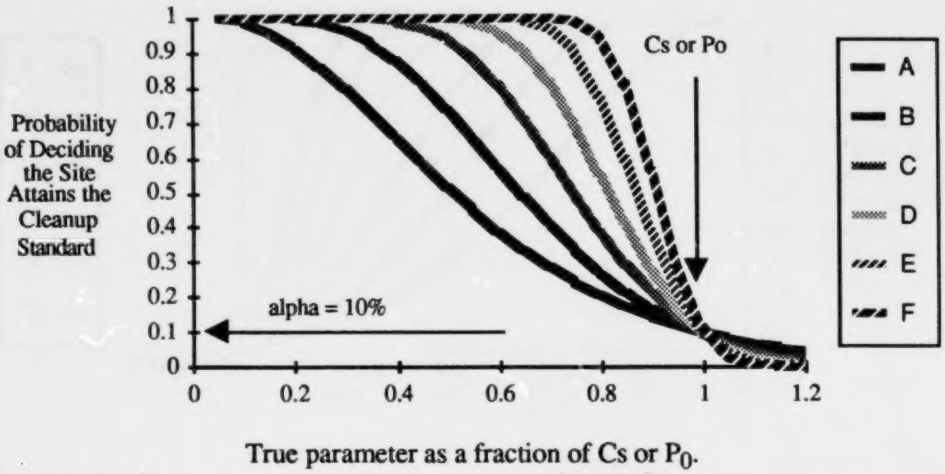
Parameters for the Power Curves	Power Curve: A	B	C	D	E	F
α =	.05	.05	.05	.05	.05	.05
β =	.20	.20	.20	.20	.20	.20
μ_1 =	.25*Cs	.43*Cs	.57*Cs	.69*Cs	.77*Cs	.84*Cs
P_1 =	.25*P_0	.43*P_0	.57*P_0	.69*P_0	.77*P_0	.84*P_0

Approximate sample sizes for simple random sampling for testing the parameters indicated

Parameters being tested	Power Curve: A	B	C	D	E	F
Mean						
with cv(data) = .5	4	5	9	17	30	61
with cv(data) = 1	11	20	34	65	117	242
with cv(data) = 1.5	25	43	76	145	264	544
Proportions						
$P_0 = 10\%$						
Non-parametric test	70	136	257	520	975	2065
Tolerance Intervals	14	33	69	151	296	649
$P_0 = 20\%$						
Non-parametric test	32	62	116	234	438	925
Tolerance Intervals	10	23	47	100	193	420

Note: α = saying the site is clean when dirty, β = saying the site is dirty when clean, $1-\beta$ = saying the site is clean when clean.

Figure A.3 Power Curves for α = 10%



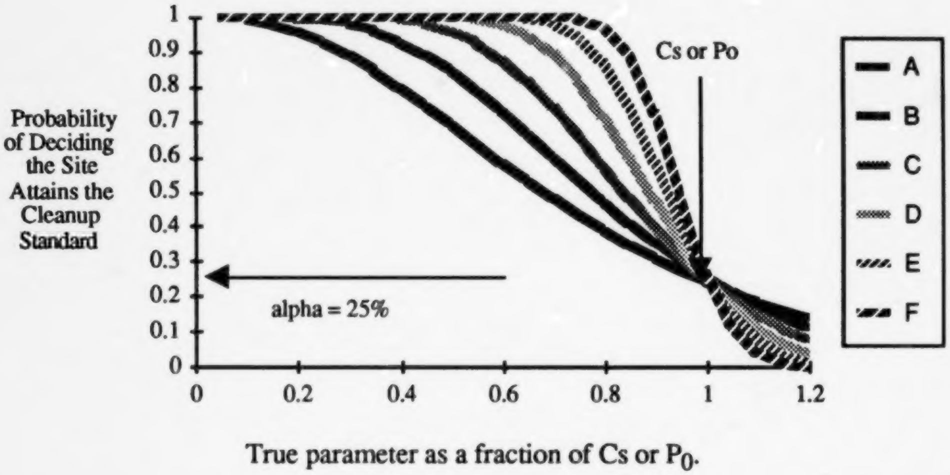
Parameters for the Power Curves	A	B	C	D	E	F
	Power Curve:					
α =	.10	.10	.10	.10	.10	.10
β =	.20	.20	.20	.20	.20	.20
μ_1 =	.30*Cs	.46*Cs	.60*Cs	.71*Cs	.79*Cs	.85*Cs
P_1 =	.30*P_0	.46*P_0	.60*P_0	.71*P_0	.79*P_0	.85*P_0

Approximate sample sizes for simple random sampling for testing the parameters indicated

Parameters being tested	A	B	C	D	E	F
	Power Curve:					
Mean						
with cv(data) = .5	3	4	8	14	26	51
with cv(data) = 1	10	16	29	54	103	201
with cv(data) = 1.5	21	35	64	121	231	452
Proportions						
P_0 = 10%						
Non-parametric test	57	108	214	430	849	1706
Tolerance Intervals	13	28	60	129	264	544
P_0 = 20%						
Non-parametric test	26	50	97	194	382	764
Tolerance Intervals	9	19	40	85	172	351

Note: α = saying the site is clean when dirty, β = saying the site is dirty when clean, 1-β = saying the site is clean when clean.

Figure A.4 Power Curves for $\alpha = 25\%$



Parameters for the	Power Curve:					
Power Curves	A	B	C	D	E	F
α =	.25	.25	.25	.25	.25	.25
β =	.20	.20	.20	.20	.20	.20
μ_1 =	.19*Cs	.40*Cs	.54*Cs	.76*Cs	.83*Cs	.87*Cs
P_1 =	$.19*P_0$	$.40*P_0$	$.54*P_0$	$.76*P_0$	$.83*P_0$	$.87*P_0$

Approximate sample sizes for simple random sampling for testing the parameters indicated

	Power Curve:					
Parameters being tested	A	B	C	D	E	F
Mean						
with cv(data) = .5	2	3	5	10	20	34
with cv(data) = 1	7	11	20	40	80	136
with cv(data) = 1.5	15	25	45	90	179	306
Proportions						
P_0 = 10%						
Non-parametric test	38	73	147	315	654	1143
Tolerance Intervals	11	22	46	100	212	375
P_0 = 20%						
Non-parametric test	18	34	67	142	294	513
Tolerance Intervals	8	15	30	66	138	242

Note: α = saying the site is clean when dirty, β = saying the site is dirty when clean, $1-\beta$ = saying the site is clean when clean.

APPENDIX B: EXAMPLE WORKSHEETS

The worksheets in this appendix have been completed to serve as an example in understanding the forms and making the necessary calculations.

The numbers and situations represented on the worksheets are hypothetical. The example situation consists of a waste site that is divided into two sample areas. The first uses random sampling to test the mean and proportion of contaminated soil for two chemicals. The second uses stratified sampling to test the mean and proportion of contaminated soil for one chemical. In this example, the different chemicals, labeled only Chemical #1, #2, and #3, are tested in the different sample areas; in most applications, the same chemicals will be tested in all or most of the sample areas. Two statistical parameters are tested for two chemicals to show how to complete the worksheet under a variety of conditions.

The following figures show the 1) the parameters being tested, 2) a hypothetical map of the site, and 3) the sequence in which the worksheets are completed. The worksheets for sample area #2 follow those for sample area #1 in this appendix.

In actual use, these worksheets would be accompanied by additional documentation such as maps, background material, justification of different choices, field notes, and copies of the results as reported by the laboratory.

APPENDIX B: EXAMPLE WORKSHEETS

Figure B.1 Example Worksheets: Parameters to Test in Each Sample Area and Map of the Site

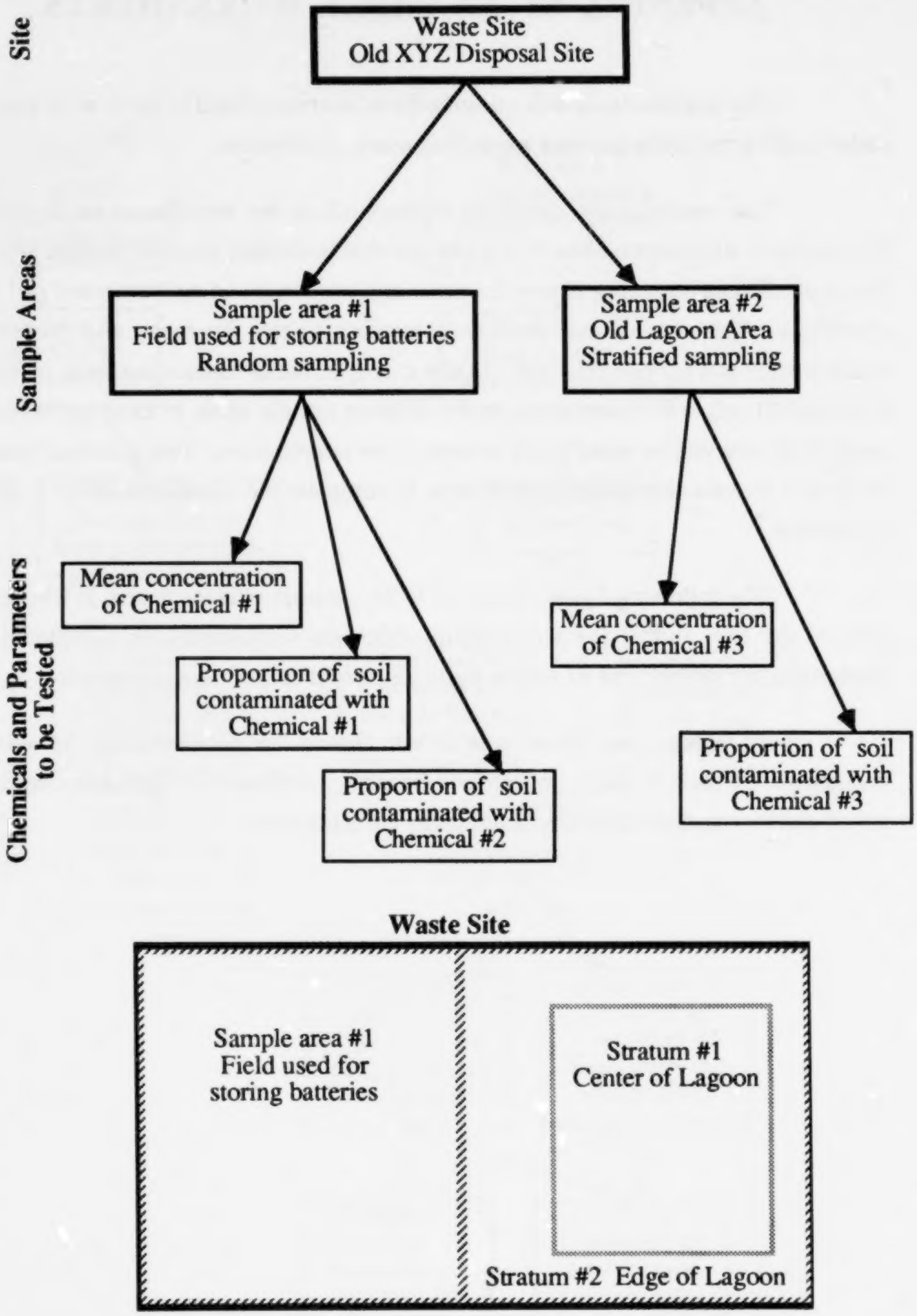
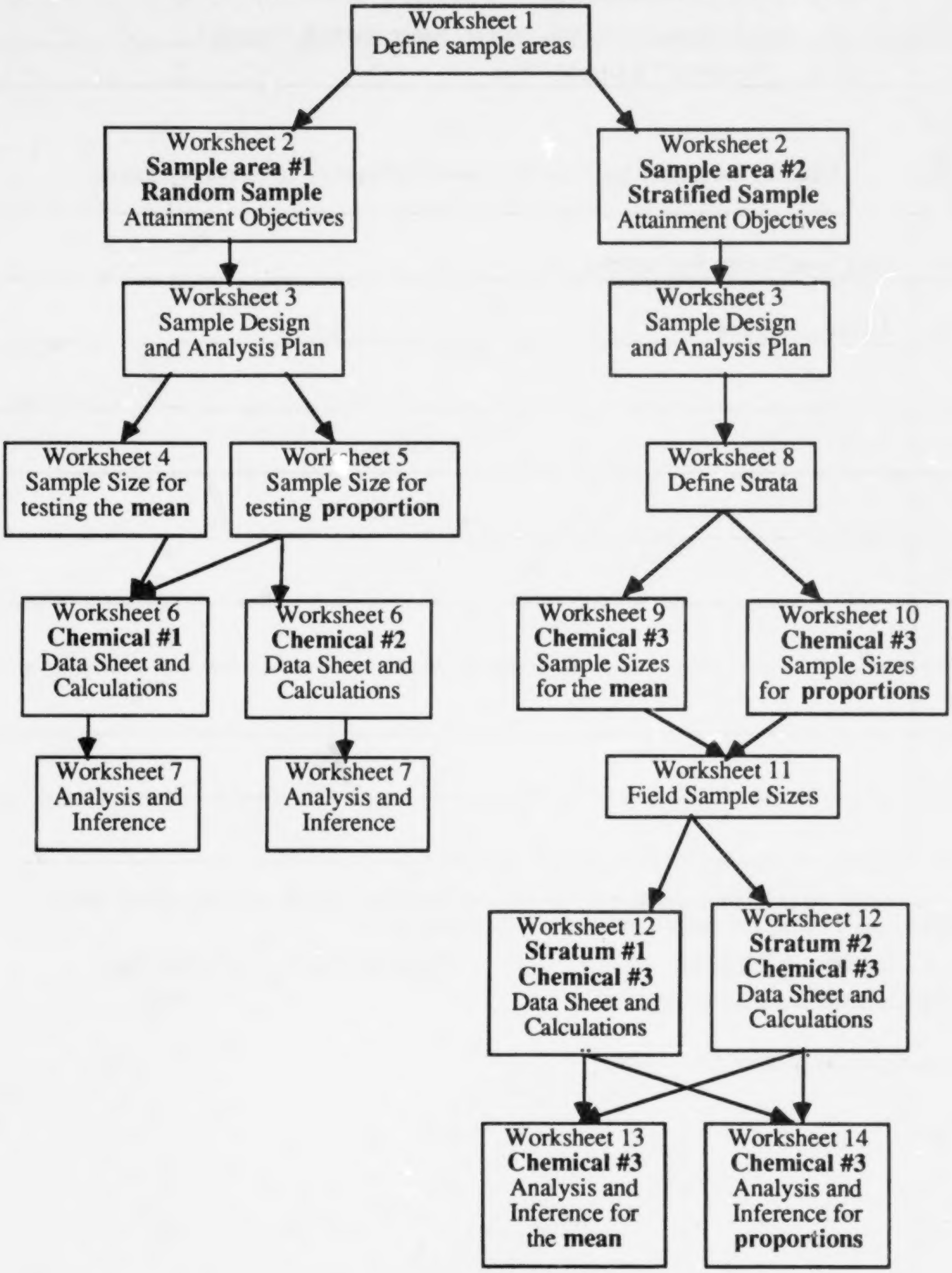

Site
Waste Site
Old XYZ Disposal Site
Sample Areas
Sample area #1
Field used for storing batteries
Random sampling
Sample area #2
Old Lagoon Area
Stratified sampling
Chemicals and Parameters to be Tested
Mean concentration of Chemical #1
Proportion of soil contaminated with Chemical #1
Proportion of soil contaminated with Chemical #2
Mean concentration of Chemical #3
Proportion of soil contaminated with Chemical #3
Waste Site
Sample area #1
Field used for storing batteries
Stratum #1
Center of Lagoon
Stratum #2 Edge of Lagoon

Figure B.2 Example Worksheets: Sequence in Which the Worksheets Are Completed


Worksheet 1
Define sample areas
Worksheet 2
Sample area #1
Random Sample
Attainment Objectives
Worksheet 2
Sample area #2
Stratified Sample
Attainment Objectives
Worksheet 3
Sample Design
and Analysis Plan
Worksheet 3
Sample Design
and Analysis Plan
Worksheet 4
Sample Size for
testing the mean
Worksheet 5
Sample Size for
testing proportion
Worksheet 8
Define Strata
Worksheet 6
Chemical #1
Data Sheet and
Calculations
Worksheet 6
Chemical #2
Data Sheet and
Calculations
Worksheet 9
Chemical #3
Sample Sizes
for the mean
Worksheet 10
Chemical #3
Sample Sizes
for proportions
Worksheet 7
Analysis and
Inference
Worksheet 7
Analysis and
Inference
Worksheet 11
Field Sample Sizes
Worksheet 12
Stratum #1
Chemical #3
Data Sheet and
Calculations
Worksheet 12
Stratum #2
Chemical #3
Data Sheet and
Calculations
Worksheet 13
Chemical #3
Analysis and
Inference for
the mean
Worksheet 14
Chemical #3
Analysis and
Inference for
proportions

WORKSHEET 1 Sample Areas

See Section 3.1 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE: Former XYZ Disposal Site

Sample Area Number g	Describe the sample areas and the reasons for treating each area separately.
1	Field used for storing batteries
2	Old Lagoon Area

Use the Sample Area Number (g) to refer on other sheets to the sample areas described above. Attach a map showing the sample areas within the waste site.

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 2**

WORKSHEET 2 Attainment Objectives

See Section 3.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]	1. Field used for storing batteries

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Collection Procedures to be used (attach separate sheet if necessary):

For example: .5 liter scoop of soil from the top 5 cm of soil, etc.

Probability of mistakenly declaring the site **clean** = α = .05

Chemical to be tested Number j	Chemical Name	Cleanup Standard (with units) Cs	Parameter to test: Mean Yes/No	Parameter to test: Proportion P_0
1	Chemical #1	20	Yes	25%
2	Chemical #2	2	No	50%

Secondary Objectives/ Other purposes for which the data is to be collected:

Use the Chemical Number (j) to refer on other sheets to the chemical described above.
Attach documentation describing the lab analysis procedure for each chemical.

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 3**

WORKSHEET 3 Sampling Design and Analysis Plan

See Chapter 4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE: Former XYZ Disposal Site

SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1] 1. Field used for storing batteries

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design:

- ☒ Simple Random Sample
- ☐ Systematic Random Sample
- ☐ Stratified Sample

Chemical to be tested Number [2] j	Comments on the Sample Design and Analysis Plan	Prob of Type II error Chance of concluding the site is dirty when it is clean β	Alternate Parameter value for the specified β Mean μ_1	Proportion P_1
1		.20	15	5%
2		.20		20%

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Continue to **WORKSHEET 4** for random or systematic sampling and **WORKSHEET 8** for stratified sampling.

WORKSHEET 4 Sample Size for Testing the Mean Using Simple Random Sampling

See Section 6.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1
If the mean concentration is not to be tested for this chemical, continue to **WORKSHEET 5**

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 1. Field used for storing batteries

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Probability of mistakenly declaring the site **clean** [2] = α: .05

From z -Table, Appendix A $z_{1-\alpha}$ = 1.645

Chemical Number [2]	[3]	From z table Appendix A	[2]	[3]		Calculate:	
j	β	$z_{1-\beta}$	Cs	μ_1	$\hat{\sigma}^2$	$A=\left(\frac{Cs-\mu_1}{z_{1-\alpha}+z_{1-\beta}}\right)^2$	$n_j=\frac{\hat{\sigma}^2}{A}$
1	.20	.842	20	15	49	4.042	12.12

Column Maximum, **Max** n_j = 12.12

Fraction of samples expected to be analyzable = **R** = .95

$\frac{\text{Max } n_j}{R}$ = **B** = 12.76

B rounded up = **Sample Size for Testing Means** = n_f = 13

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 5**

WORKSHEET 5 Sample Size for Testing Proportions Using Simple Random Sampling

See Section 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1
If the mean concentration is not to be tested for this chemical, continue to **WORKSHEET 6**

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 1. Field used for storing batteries

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From z -Table, Appendix A

Probability of mistakenly declaring the site **clean** [2] = α = .05 $z_{1-\alpha}$ = 1.645

Chemical Number [2]	[3]	From z table	[2]	[3]	Calculate:		
j	β	$z_{1-\beta}$	P_0	P_1	$A = z_{1-\alpha}\sqrt{P_0(1-P_0)}$	$B = z_{1-\beta}\sqrt{P_1(1-P_1)}$	$n_j = \left(\frac{A+B}{P_0-P_1}\right)^2$
1	.20	.842	.25	.05	.712	.184	20.06
2	.20	.842	.50	.20	.823	.337	14.93

Column Maximum, **Max** n_j = 20.06

Fraction of samples expected to be collectible = **B** = .95

$\frac{\text{Max } nj}{B}$ = **C** = 21.12

C rounded up = **Sample Size for Testing Proportions** = 22

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 6**

WORKSHEET 6 Data Calculations for a Simple Random Sample, by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]	1. Field used for storing batteries
CHEMICAL: NUMBER(j) AND DESCRIPTION [2]	1. Chemical #1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Maximum Sample Size from Worksheets 4 and 5 = Sample Size =	22
Cleanup standard[2] = Cs	20
Method Detection Limit: =	4
Concentration used when it is reported as less than the method detection limit =	4

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
1	2243	1	14.7	14.7	0	216.09
2	2244	1	17.7	17.7	0	313.29
3	2245	1	22.8	22.8	1	519.84
4	2246	1	2.9	4	0	16
5	2247	1	35.5	35.5	1	1260.25
6	2248	1	28.6	28.6	1	817.96
7	2249	1	4.9	4.9	0	24.01
8	2250	0	#N/A	0	0	0
9	2251	1	5.2	5.2	0	27.04
10	2252	1	17.2	17.2	0	295.84
Total from previous page						
Column Totals:		A 9		B 150.6	C 3	D 3490.32
		$A = n$		$B = \sum x_i$	$C = r$	$D = \sum (x_i)^2$

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 6** for other chemicals or continue to **WORKSHEET 7**

WORKSHEET 6 Data Calculations for a Simple Random Sample, by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 1. Field used for storing batteries
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2] 1. Chemical #1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Maximum Sample Size from Worksheets 4 and 5 = Sample Size =	22
Cleanup standard[2] = Cs	20
Method Detection Limit: =	4
Concentration used when it is reported as less than the detection limit =	4

Sample Number 1	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
1	2243	1	14.7	14.7	0	216.09
11	2253	1	10.9	10.9	0	118.81
12	2254	1	7.7	7.7	0	59.29
13	2255	1	12.4	12.4	0	153.76
14	2256	1	15.2	15.2	0	231.04
15	2257	1	14.9	14.9	0	222.01
16	2258	1	10.2	10.2	0	104.04
17	2259	1	17.4	17.4	0	302.76
18	2260	1	11.6	11.6	0	134.56
19	2261	1	12.4	12.4	0	153.76
20	2262	1	19.1	19.1	0	364.81
Total from previous page		9		150.6	3	3490.3
Column Totals:		A 19		B 282.4	C 3	D 5335.2
		A = n		$B = \sum x_i$	C = r	$D = \sum (x_i)^2$

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Complete **WORKSHEET 6** for other chemicals or continue to **WORKSHEET 7**

WORKSHEET 6 Data Calculations for a Simple Random Sample, by Chemical

See Section 6.3 or 7.3 in "Statistical Methods for Evaluating the Attainment of Superfund Cleanup Standards", Volume 1

SITE:		Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]	1. Field used for storing batteries
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]	1. Chemical #1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Maximum Sample Size from Worksheets 4 and 5 = Sample Size =	22
Cleanup standard [2] = Cs	20
Method Detection Limit: =	4
Concentration used when it is reported as less than the detection limit =	4

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
21	2263	1	8.9	8.9	0	79.21
22	2264	1	16.5	16.5	0	272.25
Total from previous page		19		282.4	3	5335.2
Column Totals:		A 21		B 307.8	C 3	D 5686.6
		$A = n$		$B = \sum x_i$	$C = r$	$D = \sum (x_i)^2$

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 6** for other chemicals or continue to **WORKSHEET 7**

WORKSHEET 7 Inference for Simple Random Samples by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:		Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]	1. Field used for storing batteries
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]	1. Chemical #1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Testing the Mean

[2] $\alpha =$ .05

[2] $Cs =$ 20

Number of Collectible Samples [6] = $\mathbf{n} =$ 21

Total of the concentration measurements [6] = $\sum x_i = \mathbf{B} =$ 307.8

Total for x_i^2 [6] = $\sum (x_i)^2 = \mathbf{D} =$ 5687

Mean concentration = $\frac{B}{n} = \bar{x} =$ 14.66

Standard Deviation of the Data = $\sqrt{\frac{D - n\bar{x}^2}{n-1}} = \mathbf{s} =$ 7.67

Degrees of Freedom for s = n - 1 = **df** = 20

$t_{1-\alpha,df} =$ 1.73

Standard Error for the Mean concentration = $\frac{s}{\sqrt{n}} =$ 1.67

Upper One Sided Confidence Interval = $\bar{x} + t_{1-\alpha,df}\frac{s}{\sqrt{n}} = \mu_{U\alpha} =$ 17.54

If $\mu_{U\alpha} < Cs$ then circle Clean, otherwise circle Dirty:
Based on the mean concentration, the sample area is: ***Clean*** ***Dirty***

Testing Percentiles

[2] $P_0 =$ .25

[4 or 5] $z_{1-\alpha} =$ 1.645

Number of Samples with Concentration Greater than Cs [6] = $\mathbf{r} =$ 3

Proportion of Contaminated Samples = $\frac{r}{n} = \mathbf{p} =$ .143

Standard Error for the Proportion = $\sqrt{\frac{p(1-p)}{n}} = s_p =$ .0764

Test Statistic = $p + z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}} = U_L =$ .298

If $U_L < P_0$ then circle Clean, otherwise circle Dirty:
Based on the proportion of contaminated samples, the sample area is: ***Clean*** ***Dirty***

Date Completed: ***EXAMPLE*** Completed by ***EXAMPLE***

Page ____ of ____

Complete **WORKSHEET 7** for other chemicals

WORKSHEET 6 Data Calculations for a Simple Random Sample, by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site	
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]	1. Field used for storing batteries
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]	2. Chemical #2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Maximum Sample Size from Worksheets 4 and 5 = Sample Size =	22
Cleanup standard [2] = Cs	20
Method Detection Limit: =	1.2
Concentration used when it is reported as less than the detection limit =	1.2

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
1		1	1.2	1.2	0	
2		1	2.1	2.1	1	
3		1	0.9	1.2	0	
4		1	0.1	1.2	0	
5		1	0.5	1.2	0	
6		1	0.3	1.2	0	
7		1	0.3	1.2	0	
8		0	#N/A	0	0	
9		1	1.9	1.9	0	
10		1	8.3	8.3	1	
Total from previous page						
Column Totals:		A 9		B 19.5	C 2	D
		A = n		$B = \sum x_i$	C = r	$D = \sum (x_i)^2$

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 6** for other chemicals or continue to **WORKSHEET 7**

WORKSHEET 6 Data Calculations for a Simple Random Sample, by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 1. Field used for storing batteries
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2] 2. Chemical #2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Maximum Sample Size from Worksheets 4 and 5 = Sample Size =	22
Cleanup standard [2] = Cs	20
Method Detection Limit: =	1.2
Concentration used when it is reported as less than the detection limit =	1.2

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
11		1	0.5	1.2	0	
12		1	0.7	1.2	0	
13		1	2.2	2.2	1	
14		1	0.7	1.2	0	
15		1	1.7	1.7	0	
16		1	2.3	2.3	1	
17		1	0.3	1.2	0	
18		1	3.7	3.7	1	
19		1	0.1	1.2	0	
20		1	5.6	5.6	1	
Total from previous page		9		19.5	2	
Column Totals:		A 19		B 41	C 6	D
		$A = n$		$B = \sum x_i$	$C = r$	$D = \sum (x_i)^2$

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 6** for other chemicals or continue to **WORKSHEET 7**

WORKSHEET 6 Data Calculations for a Simple Random Sample, by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]	1. Field used for storing batteries
CHEMICAL: NUMBER(j) AND DESCRIPTION [2]	2. Chemical #2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Maximum Sample Size from Worksheets 4 and 5 = Sample Size =	22
Cleanup standard[2] = Cs	20
Method Detection Limit: =	4
Concentration used when it is reported as less than the detection limit =	4

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
21		1	1.3	1.3	0	
22		1	1.8	1.8	0	
Total from previous page		19		41	6	
Column Totals:		A 21		B 44.1	C 6	D
		$A = n$		$B = \sum x_i$	$C = r$	$D = \sum (x_i)^2$

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Complete **WORKSHEET 6** for other chemicals or continue to **WORKSHEET 7**

WORKSHEET 7 Inference for Simple Random Samples by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 1. Field used for storing batteries
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2] 2. Chemical #2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Testing the Mean

[2] α =	.05
[2] Cs =	2
Number of Collectible Samples [6] = **n** =	21
Total of the concentration measurements [6] = $\sum x_i$ = **B** =	44.1
Total for x_i^2 [6] = $\sum (x_i)^2$ = **D** =	
Mean concentration = $\frac{B}{n}$ = $\bar{x}$ =	
Standard Deviation of the Data = $\sqrt{\frac{D-n\bar{x}^2}{n-1}}$ = **s** =	
Degrees of Freedom for s = n - 1 = **df** =	
$t_{1-\alpha,df}$ =	
Standard Error for the Mean concentration = $\frac{s}{\sqrt{n}}$ =	
Upper One Sided Confidence Interval = $\bar{x} + t_{1-\alpha,df}\, s_\mu \sqrt{n}) = \mu_{U\alpha}$ =	
If $\mu_{U\alpha} <$ Cs then circle Clean, otherwise circle Dirty: Based on the mean concentration, the sample area is:	***Clean Dirty***

Testing Percentiles

[2] P_0 =	.5
[4 or 5] $z_{1-\alpha}$ =	1.645
Number of Samples with Concentration Greater than Cs [6] = **r** =	6
Proportion of Contaminated Samples = $\frac{r}{n}$ = **p** =	.286
Standard Error for the Proportion = $\sqrt{\frac{p(1-p)}{n}} = s_p$ =	.0986
Test Statistic = $p + z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}} = U_L$ =	.465
If $U_L < P_0$ then circle Clean, otherwise circle Dirty: Based on the proportion of contaminated samples, the sample area is:	***Clean Dirty***

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Page ____ of ____

Complete **WORKSHEET 7** for other chemicals

WORKSHEET 2 Attainment Objectives

See Section 3.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE: Former XYZ Disposal Site

SAMPLE AREA: NUMBER(s) AND DESCRIPTION [1] 1. Old Lagoon Area

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Collection Procedures to be used (attach separate sheet if necessary):

For example: One foot soil core, 2 inches in diameter, etc.

Probability of mistakenly declaring the site **clean** = α = .05

Chemical to be tested Number j	Chemical Name	Cleanup Standard (with units) Cs	Parameter to test:	
			Mean Yes/No	Proportion P_0
1	Chemical #3	30	Yes	25%

Secondary Objectives/ Other purposes for which the data is to be collected:

Use the Chemical Number (j) to refer on other sheets to the chemical described above.
Attach documentation describing the lab analysis procedure for each chemical.

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Continue to **WORKSHEET 3**

WORKSHEET 3 Sampling Design and Analysis Plan

See Chapter 4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 1. Old Lagoon Area

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design:

☐ Simple Random Sample
☐ Systematic Random Sample
☒ Stratified Sample

Chemical to be tested Number [2] j	Comments on the Sample Design and Analysis Plan	Prob of Type II error Chance of concluding the site is dirty when it is clean β	Alternate Parameter value for the specified β Mean μ_1	Alternate Parameter value for the specified β Proportion P_1
3	Use non-parametric estimation of	.20	15	10%
	proportions			

Date Completed: *EXAMPLE*　　Completed by *EXAMPLE*

Use additional sheets if necessary.　　Page ____ of ____

Continue to WORKSHEET 4 for random or systematic sampling and **WORKSHEET 8** for stratified sampling.

WORKSHEET 8 Definition of Strata Within Sample Area

See Section 4.1 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 2. Old Lagoon Area

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Stratum Number h	Describe the stratum and the reason for interest in this area	Volume = Surface Area * Sample depth V_h	$W_h = \frac{V_h}{\Sigma V_h}$
1	Center of Lagoon	141,000 cu. ft.	.60
2	Edge of Lagoon	94,000 cu. ft.	.40

Total Volume = ΣV_h = 235,000 cu. ft.

Use the Stratum Number (h) to refer on other worksheets to the stratum described above
Attach a map showing the stratum within the sample area.

Date Completed: *EXAMPLE*　　Completed by *EXAMPLE*

Use additional sheets if necessary.　　Page ____ of ____

Continue to **WORKSHEET 9**

WORKSHEET 9 Desired Sample Sizes for Testing the Mean Using Stratified Sampling, by Chemical

See Section 6.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:		Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]	2. Old Lagoon Area
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]	Chemical #3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From z -Table, Appendix A

Probability of mistakenly declaring the site **clean** [2] = α = .05 $z_{1-\alpha}$ = 1.645

For the Cleanup Standard = Cs = 30

Probability of mistakenly declaring the site **dirty** [3] = β = .20 $z_{1-\beta}$ = .842

If the true concentration is [3] = μ_1 = 15

Calculate: $$\left\{\frac{Cs - \mu_1}{z_{1-\alpha} + z_{1-\beta}}\right\}^2 = A = 36.38$$

Stratum Number[8]	Proportion of Sample Area in Stratum[8]	Stratum Standard Deviation	Unit Sample Cost			Desired final sample size n_{hd} =	Calculation check
h	W_h	$\hat{\sigma}_h$	C_h	$W_h * \hat{\sigma}_h \sqrt{C_h}$	$\frac{W_h * \hat{\sigma}_h}{\sqrt{C_h}}$	$\frac{B * W_h * \hat{\sigma}_h}{\sqrt{C_h}}$	$\frac{(W_h \hat{\sigma}_h)^2}{n_h}$
1	.6	35	1	21	21	17.2	25.64
2	.4	22	1	8.8	8.9	7.21	10.74

C = Column Sum = 29.8 A = 36.38

B = C/A = B .819

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 10**

WORKSHEET 10 Desired Sample Sizes for Testing a Percentile Using Stratified Sampling, by Chemical

See Section 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 2. Old Lagoon Area
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2] 3. Chemical #3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From z -Table, Appendix A

Probability of mistakenly declaring the site **clean** [2] = α = .05 $z_{1-\alpha}$ = 1.645

Proportion Exceeding Cleanup Standard [2] = P_0 = .25

Probability of mistakenly declaring the site **dirty** [3] = β = .2 $z_{1-\beta}$ = .842

If the true proportion is [2] = P_1 = .1

Calculate: $\left\{\frac{P_0 - P_1}{z_{1-\alpha} + z_{1-\beta}}\right\}^2$ = .00364

Stratum Number[3] h	Proportion of Sample Area in Stratum[3] W_h	Proportion of dirty Samples p_h	Stratum Standard Deviation $\hat{\sigma}_h$ $\sqrt{P_h(1-P_h)}$	Unit Sample Cost C_h	$W_h{*}\hat{\sigma}_h\sqrt{C_h}$	$\frac{W_h{*}\hat{\sigma}_h}{\sqrt{C_h}}$	Desired final sample size $n_{hd} = \frac{B{*}W_h{*}\hat{\sigma}_h}{\sqrt{C_h}}$	Calculation check $\frac{(W_h\hat{\sigma}_h)^2}{n_h}$
1	.6	.145	.352	1	.211	.211	16.54	.0027
2	.4	.036	.184	1	.074	.074	6.76	.0009

C = Column Sum = .285 A = .0036

Divided by A =

B = C/A = B 78.3

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 11**

WORKSHEET 11 Desired Sample Sizes for All Chemicals and Parameters

See Section 6.4 or 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 2. Old Lagoon Area

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Enter the Desired Sample Size (n_{hd})by Stratum for each combination of chemical and parameter to be tested, from WORKSHEETS 9 and 10.

Desired Sample Size by stratum and : chemical	Stratum number h 1	2	3	4	5
Chemical 3 Worksheet 9	17.2	7.2			
Chemical 3 Worksheet 10	16.54	5.75			
Chemical __ Worksheet __					
Chemical __ Worksheet __					
Chemical __ Worksheet __					
Maximum n_{hd} for all Chemicals and Parameters n_{hdmax}	17.2	7.2			
Fraction of Collectible Field Samples R_h	.95	.95			
$A = \frac{n_{hdmax}}{R_h}$	18.1	7.6			
A Rounded up to the Next Integer = n_{hf}, the field sample size	19	8			

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Continue to **WORKSHEET 12**

WORKSHEET 12 Data Calculations, by Stratum and Chemical

See Section 6.4 or 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]	2. Old Lagoon Area
STRATUM: NUMBER AND DESCRIPTION [3]	1. Center of Lagoon
CHEMICAL: NUMBER(j) AND DESCRIPTION [2]	3. Chemical #3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Samples Size [11] = n_{hf} = 19

Cleanup standard[2] = Cs = 30

Method Detection Limit: = 5

Concentration used when no concentration is reported = 5

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
1	9301	1	17	17	0	289
2	9302	1	40	40	1	1600
3	9303	0	#NA	0	0	0
4	9304	1	19	19	0	361
5	9305	1	31	31	1	961
6	9306	1	0	5	0	25
7	9307	1	7	7	0	49
8	9308	1	10	10	0	100
9	9309	1	15	15	0	225
10	9310	1	21	21	0	441
Total from previous page						
Column Totals:		A 9		B 165	C 2	D 4051
		$n_h = A$		$\sum h_i = B$	$r_h = C$	$\sum (h_i)^2 = D$

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 12** for other chemicals or to **WORKSHEET 13**

WORKSHEET 12 Data Calculations, by Stratum and Chemical

See Section 6.4 or 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 2. Old Lagoon Area
STRATUM:	NUMBER AND DESCRIPTION [3] 1. Center of Lagoon
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2] 3. Chemical #3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Samples Size [11] = n_{hf} =	19
Cleanup standard[2] = Cs =	30
Method Detection Limit: =	5
Concentration used when no concentration is reported =	5

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
11	9311	1	18	18	0	324
12	9312	1	27	27	0	729
13	9313	1	12	12	0	144
14	9314	1	28	28	0	784
15	9315	1	94	94	1	8836
16	9316	0	#NA	0	0	0
17	9317	1	13	13	0	169
18	9318	1	22	22	0	484
19	9319	1	23	23	0	529
Total from previous page		9		165	2	405
Column Totals:		A 17		B 402	C 3	D 16050
		$n_h = A$		$\sum x_{hi} = B$	$r_h = C$	$\sum (x_{hi})^2 = D$

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 12** for other chemicals or to **WORKSHEET 13**

WORKSHEET 12 Data Calculations, by Stratum and Chemical

See Section 6.4 or 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume. 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 2. Old Lagoon Area
STRATUM:	NUMBER AND DESCRIPTION [3] 2. Edge of Lagoon
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2] 3. Chemical #3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Samples Size [11] = n_{hf} =	19
Cleanup standard[2] = Cs =	30
Method Detection Limit: =	5
Concentration used when no concentration is reported =	5

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
1	9320	1	9	9	0	81
2	9321	1	16	16	0	256
3	9322	0	#NA	0	0	0
4	9323	1	0	5	0	25
5	9324	1	0	5	0	25
6	9325	1	5	5	0	25
7	9326	1	8	8	0	64
8	9327	1	10	10	0	100
9	9328	1	9	9	0	81
Total from previous page						
Column Totals:		A 8		B 67	C 0	D 657
		$n_h = A$		$\sum h_i = B$	$r_h = C$	$\sum (h_i)^2 = D$

Date Completed: EXAMPLE

Completed by EXAMPLE

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 12** for other chemicals or to **WORKSHEET 13**

WORKSHEET 13 Sample Area Analysis for the Mean Using Stratified Sampling, by Chemical

See Section 6.4 in " Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 2. Old Lagoon Area
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2] 3. Chemical #3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Stratum Number[3] h	[3] W_h	[12] $n_h = A$	[12] $\bar{x}_h = \frac{B}{A}$	[12] $s_h^2 = \frac{D-\bar{x}_h^2 A}{A-1}$	$W_h\bar{x}_h$	$F = \frac{W_h^2 s_h^2}{n_h}$	$\frac{F^2}{(n_h-1)}$
1	0.6	17	23.65	408.99	14.19	8.661	4.6883
2	0.4	8	8.38	13.70	3.35	0.274	0.0107
Grand Totals:					G 17.54	H 8.935	I 4.6990

[2] α = .05

[2] Cs = 30

Mean concentration = G = $\bar{x}$ = 17.54

Degrees of Freedom = $\frac{H^2}{I}$ Rounded to an integer = **df** = 17

$t_{1-\alpha,df}$ = 1.74

Standard Error for the Mean concentration = $\sqrt{H} = s_{\mu}$ = 2.99

Upper One Sided Confidence Interval = $\bar{x} + s_{\mu}\, t_{1-\alpha,df} = \mu_{U\alpha}$ = 22.74

If $\mu_{U\alpha} <$ Cs then circle Clean, otherwise circle Dirty:
Based on the mean concentration, the sample area is: ***Clean*** ***Dirty***

Date Completed: *EXAMPLE* Completed by *EXAMPLE*

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 14**

WORKSHEET 14 Sample Area Analysis for a Percentile Using Stratified Sampling, by Chemical

See Section 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	Former XYZ Disposal Site
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1] 2. Old Lagoon Area
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Stratum Number[3] h	[3] W_h	[9] n_h	[9] $p_h = \frac{C}{A}$	[9] $s_h^2 = p_h(1-p_h)$	$p = W_h p_h$	$F = W_h^2 \frac{s_h^2}{n_h}$
1	0.6	17	0.18	0.1453	0.1059	0.00308
2	0.4	8	0.00	0	0	0

Grand Totals:

G 0.1059	H .00308

[2] P_0 = .25

[4 or 5] $z_{1-\alpha}$ = 1.645

Proportion of Contaminated Samples = G = p = .106

Standard Error for the Proportion = $\sqrt{H} = s_p$ = .055

Test Statistic = $p + s_p z_{1-\alpha} = T$ = .197

If $T < P_0$ then circle Clean, otherwise circle Dirty:
Based on the mean concentration, the sample area is: ***Clean Dirty***

Date Completed: *EXAMPLE*

Completed by *EXAMPLE*

Use additional sheets if necessary.

Page ____ of ____

APPENDIX C: BLANK WORKSHEETS

The worksheets in this appendix may be used or modified to document the decisions, record data, and make calculations to determine if the waste site attains the cleanup standard. These worksheets are referred to in the document. Appendix B provides examples of how to fill out the worksheets.

APPENDIX C: BLANK WORKSHEETS

WORKSHEET 1 Sample Areas

See Section 3.1 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:

Sample Area Number g	Describe the sample areas and the reasons for treating each area separately.

Use the Sample Area Number (g) to refer on other sheets to the sample areas described above. Attach a map showing the sample areas within the waste site.

Date Completed: ____________ Completed by ___________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 2**

WORKSHEET 2 Attainment Objectives

See Section 3.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:

SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Collection Procedures to be used (attach separate sheet if necessary):

Probability of mistakenly declaring the site **clean** = α = ☐

Chemical to be tested Number j	Chemical Name	Cleanup Standard (with units) Cs	Parameter to test: Mean Yes/No	Proportion P_0

Secondary Objectives/ Other purposes for which the data is to be collected:

Use the Chemical Number (j) to refer on other sheets to the chemical described above.
Attach documentation describing the lab analysis procedure for each chemical.

Date Completed: ____________ Completed by ________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 3**

WORKSHEET 3 Sampling Design and Analysis Plan

See Chapter 4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:

SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design:

☐ Simple Random Sample
☐ Systematic Random Sample
☐ Stratified Sample

Chemical to be tested Number [2] j	Comments on the Sample Design and Analysis Plan	Prob of Type II error Chance of concluding the site is dirty when it is clean β	Alternate Parameter value for the specified β Mean μ_1	Alternate Parameter value for the specified β Proportion P_1

Date Completed: ____________ Completed by ________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 4** for random or systematic sampling and **WORKSHEET 8** for stratified sampling.

WORKSHEET 4 Sample Size for Testing the Mean Using Simple Random Sampling

See Section 6.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

If the mean concentration is not to be tested for this chemical, continue to **WORKSHEET 5**

SITE:	
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Probability of mistakenly declaring the site **clean** [2] = α ☐

From z -Table, Appendix A $z_{1-\alpha} =$ ☐

Chemical Number [2]	[3]	From z table Appendix A	[2]	[3]		Calculate:	
j	β	$z_{1-\beta}$	Cs	μ_1	$\hat{\sigma}^2$	$A=\left(\frac{Cs-\mu_1}{z_{1-\alpha}+z_{1-\beta}}\right)^2$	$n_j = \frac{\hat{\sigma}^2}{A}$

Column Maximum, **Max n_j** = ☐

Fraction of samples expected to be analyzable = **R** = ☐

$\frac{\text{Max } nj}{R}$ = **B** = ☐

B rounded up = **Sample Size for Testing Means** = n_f = ☐

Date Completed: ____________

Completed by ________________________

Use additional sheets if necessary.

Page ____ of ____

Continue to **WORKSHEET 5**

WORKSHEET 5 Sample Size for Testing Proportions Using Simple Random Sampling

See Section 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1
If the mean concentration is not to be tested for this chemical, continue to **WORKSHEET 6**

SITE:	
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Probability of mistakenly declaring the site **clean** [2] = α

From z -Table, Appendix A
$z_{1-\alpha}$ =

Chemical Number [2] j	[3] β	From z table $z_{1-\beta}$	[2] P_0	[3] P_1	Calculate: $A = z_{1-\alpha}\sqrt{P_0(1-P_0)}$	$B = z_{1-\beta}\sqrt{P_1(1-P_1)}$	$n_j = \left(\frac{A+B}{P_0-P_1}\right)^2$

Column Maximum, **Max n_j** =

Fraction of samples expected to be collectible = **B** =

$\frac{\text{Max } nj}{B}$ = **C** =

C rounded up to the next integer = **Sample Size for Testing Proportions** =

Date Completed: ____________ Completed by ________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 6**

WORKSHEET 6 Data Calculations for a Simple Random Sample, by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Maximum Sample Size from Worksheets 4 and 5 = Sample Size =	
Cleanup standard [2] = Cs	
Method Detection Limit: =	
Concentration used when it is reported as less than the method detection limit=	

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
Total from previous page						
Column Totals:		A		B	C	D
		A = n		$B = \sum x_i$	C = r	$D = \sum (x_i)^2$

Date Completed: ____________

Completed by ________________________

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 6** for other chemicals or continue to **WORKSHEET 7**

WORKSHEET 7 Inference for Simple Random Samples by Chemical

See Section 6.3 or 7.3 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Testing the Mean

[2] $\alpha =$

[2] $Cs =$

Number of Collectible Samples [6] = $\mathbf{n} =$

Total of the concentration measurements [6] = $\sum x_i = \mathbf{B} =$

Total for x_i^2 [6] = $\sum (x_i)^2 = \mathbf{D} =$

Mean concentration $= \frac{B}{n} = \bar{x} =$

Standard Deviation of the Data $= \sqrt{\frac{D - n\bar{x}^2}{n-1}} = s =$

Degrees of Freedom for s = n - 1 = **df** =

$t_{1-\alpha,df} =$

Standard Error for the Mean concentration $= \frac{s}{\sqrt{n}} =$

Upper One Sided Confidence Interval $= \bar{x} + t_{1-\alpha,df} \frac{s}{\sqrt{n}} = \mu_{U\alpha} =$

If $\mu_{U\alpha} < Cs$ then circle Clean, otherwise circle Dirty:
Based on the mean concentration, the sample area is: ***Clean Dirty***

Testing Percentiles

[2] $P_0 =$

[4 or 5] $z_{1-\alpha} =$

Number of Samples with Concentration Greater than Cs [6] = $\mathbf{r} =$

Proportion of Contaminated Samples $= \frac{r}{n} = p =$

Standard Error for the Proportion $= \sqrt{\frac{p(1-p)}{n}} = s_p =$

Test Statistic $= \frac{p - p_0}{\sqrt{\frac{p(1-p)}{n}}} = z =$

If $z < z_{1-\alpha}$ then circle Clean, otherwise circle Dirty:
Based on the proportion of contaminated samples, the sample area is: ***Clean Dirty***

Date Completed: ____________ Completed by ________________________

Page ____ of ____

Complete **WORKSHEET 7** for other chemicals

WORKSHEET 8 Definition of Strata Within Sample Area

See Section 4.1 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:

SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Stratum Number h	Describe the stratum and the reason for interest in this area	Volume = Surface Area * Sample depth V_h	$W_h = \frac{V_h}{\Sigma V_h}$

Total Volume = ΣV_h =

Use the Stratum Number (h) to refer on other worksheets to the stratum described above
Attach a map showing the stratum within the sample area.

Date Completed: ____________ Completed by ________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 9**

WORKSHEET 9 Desired Sample Sizes for Testing the Mean Using Stratified Sampling, by Chemical

See Section 6.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From z -Table, Appendix A

Probability of mistakenly declaring the site **clean** [2] = α ☐ $z_{1-\alpha} =$ ☐

For the Cleanup Standard = Cs = ☐

Probability of mistakenly declaring the site **dirty** [3] = β = ☐ $z_{1-\beta} =$ ☐

If the true concentration is [3] = μ_1 = ☐

Calculate: $\left\{\dfrac{Cs - \mu_1}{z_{1-\alpha} + z_{1-\beta}}\right\}^2 = A =$ ☐

Stratum Number[8] h	Proportion of Sample Area in Stratum[8] W_h	Stratum Standard Deviation $\hat{\sigma}_h$	Unit Sample Cost C_h	$W_h * \hat{\sigma}_h \sqrt{C_h}$	$\dfrac{W_h * \hat{\sigma}_h}{\sqrt{C_h}}$	Desired final sample size $n_{hd} = \dfrac{B * W_h * \hat{\sigma}_h}{\sqrt{C_h}}$	Calculation check $\dfrac{(W_h \hat{\sigma}_h)^2}{n_h}$

C = Column Sum = ☐

A = ☐

B = C/A = B

Date Completed: ____________ Completed by ________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 10**

WORKSHEET 10 Desired Sample Sizes for Testing a Percentile Using Stratified Sampling, by Chemical

See Section 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From z -Table, Appendix A

Probability of mistakenly declaring the site **clean** [2] = α ☐ $z_{1-\alpha} =$ ☐

Proportion Exceeding Cleanup Standard [2] = $P_0 =$ ☐

Probability of mistakenly declaring the site **dirty** [3] = $\beta =$ ☐ $z_{1-\beta} =$ ☐

If the true proportion is [2] = $P_1 =$ ☐

Calculate: $\left\{\frac{P_0 - P_1}{z_{1-\alpha} + z_{1-\beta}}\right\}^2 = A =$ ☐

Stratum Number[3] h	Proportion of Sample Area in Stratum[3] W_h	Proportion of dirty Samples p_h	Stratum Standard Deviation $\hat{\sigma}_h$ $\sqrt{P_h(1-P_h)}$	Unit Sample Cost C_h	$W_h{*}\hat{\sigma}_h\sqrt{C_h}$	$\frac{W_h{*}\hat{\sigma}_h}{\sqrt{C_h}}$	Desired final sample size $n_{hd} = \frac{B{*}W_h{*}\hat{\sigma}_h}{\sqrt{C_h}}$	Calculation check $\frac{(W_h\hat{\sigma}_h)^2}{n_h}$

C = Column Sum = ☐ A = ☐

B = C/A = B

Date Completed: ____________ Completed by ________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 11**

WORKSHEET 11 Desired Sample Sizes for All Chemicals and Parameters

See Section 6.4 or 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:

SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Enter the Desired Sample Size (n_{hd})by Stratum for each combination of chemical and parameter to be tested, from WORKSHEETS 9 and 10.

Desired Sample Size by stratum and : chemical	Stratum number h 1	2	3	4	5
Chemical __ Worksheet __					
Chemical __ Worksheet __					
Chemical __ Worksheet __					
Chemical __ Worksheet __					
Chemical __ Worksheet __					
Maximum n_{hd} for all Chemicals and Parameters n_{hdmax}					
Fraction of Collectible Field Samples R_h					
$A = \frac{n_{hdmax}}{R_h}$					
A Rounded up to the Next Integer = n_{hf}, the field sample size					

Date Completed: ____________ Completed by ________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 12**

WORKSHEET 12 Data Calculations, by Stratum and Chemical

See Section 6.4 or 7.4 in " Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:

SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]

STRATUM: NUMBER AND DESCRIPTION [3]

CHEMICAL: NUMBER(j) AND DESCRIPTION [2]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Samples Size [11] = n_{hf} =

Cleanup standard[2] = Cs =

Method Detection Limit: =

Concentration used when no concentration is reported =

Sample Number i	Sample ID	Was the Sample Collectible? 0 = No 1 = Yes	Reported Concentration If Collectible	Concentration Corrected for Detection Limit x_i	Is x_i Greater than Cs? 1 = Yes 0 = No y_i	$(x_i)^2$
Total from previous page						
Column Totals:		A		B	C	D
		$n_h = A$		$\sum h_i = B$	$r_h = C$	$\sum (h_i)^2 = D$

Date Completed: ____________

Completed by ________________________

Use additional sheets if necessary.

Page ____ of ____

Complete **WORKSHEET 12** for other chemicals or to **WORKSHEET 13**

WORKSHEET 13 Sample Area Analysis for the Mean Using Stratified Sampling, by Chemical

See Section 6.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:	
SAMPLE AREA:	NUMBER(g) AND DESCRIPTION [1]
CHEMICAL:	NUMBER(j) AND DESCRIPTION [2]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Stratum Number[3]	[3]	[12]	[12]	[12]			
h	W_h	$n_h = A$	$\bar{x}_h = \frac{B}{A}$	$s_h^2 = \frac{D-\bar{x}_h^2 A}{A-1}$	$W_h\bar{x}_h$	$F = \frac{W_h^2 s_h^2}{A}$	$\frac{F^2}{(A-1)}$
Grand Totals:					G	H	I

[2] $\alpha =$ ____

[2] Cs = ____

Mean concentration = G = $\bar{x}$ = ____

Degrees of Freedom = $\frac{H^2}{I}$ Rounded to an integer = **df** = ____

$t_{1-\alpha,df} =$ ____

Standard Error for the Mean concentration = $\sqrt{H} = s_\mu =$ ____

Upper One Sided Confidence Interval = $\bar{x} + s_\mu t_{1-\alpha,df} = \mu_{U\alpha} =$ ____

If $\mu_{U\alpha} <$ Cs then circle Clean, otherwise circle Dirty:
Based on the mean concentration, the sample area is: ***Clean*** ***Dirty***

Date Completed: ____________ Completed by ________________________

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 14**

WORKSHEET 14 Sample Area Analysis for a Percentile Using Stratified Sampling, by Chemical

See Section 7.4 in "Methods for Evaluating the Attainment of Cleanup Standards," Volume 1

SITE:

SAMPLE AREA: NUMBER(g) AND DESCRIPTION [1]

CHEMICAL: NUMBER(j) AND DESCRIPTION [2]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Stratum Number[3]	[3]	[9]	[9]	[9]		
h	W_h	n_h	$p_h = \frac{C}{A}$	$s_h^2 = p_h(1-p_h)$	$p = W_h p_h$	$F = W_h^2 \frac{s_h^2}{n_h}$
Grand Totals:					G	H

[2] P_0 =

[4 or 5] $z_{1-\alpha}$ =

Proportion of Contaminated Samples = G = **p** =

Standard Error for the Proportion = $\sqrt{H}$ = s_p =

Test Statistic = $p + s_p z_{1-\alpha}$ = T =

If $T < P_0$ then circle Clean, otherwise circle Dirty:
Based on the mean concentration, the sample area is: ***Clean*** ***Dirty***

Date Completed: ____________

Completed by ________________________

Use additional sheets if necessary.

Page ____ of ____

APPENDIX D: GLOSSARY

Alpha (α) In the context of a statistical test, α is probability of a Type I error.

Alternative Hypothesis See *hypothesis.*

Analysis Plan The plan specifying how the data are to be analyzed once they are collected, including what estimates are to be made from the data, how the estimates are to be calculated, and how the results of the analysis will be reported.

Attainment The achievement of a prescribed standard/level of concentration.

Attainment Objectives Specifying chemicals to be tested, specifying the cleanup standard to be attained, specifying the measure or parameter to be compared to the cleanup standard, and specifying the level of confidence required if the environment and human health are to be protected.

Beta (β) In the context of a statistical test, β is probability of a Type II error.

Binomial Distribution A probability distribution used to describe the number of occurrences of a specified event in n independent trials. In this manual, the binomial distribution is used to develop statistical tests concerned with testing the proportion of soil units in a simple random sample having excessive concentrations of a contaminant (see Chapter 7). For additional details about the binomial distribution, consult Conover (1980).

Coefficient of Variation The ratio of the standard deviation to the mean for a set of data or distribution, abbreviated **cv**. For data that can only have positive values, such as concentration measurements, the coefficient of variation provides a crude measure of skewness.

Confidence Interval A sample-based estimate of a population parameter expressed as a range or interval of values, rather than as a single value (point estimate).

Confidence Level The degree of confidence associated with an interval estimate. For example, with a 95% confidence interval, we would be 95% certain that the interval contains the true value being estimated. The confidence level is equal to 1 minus the Type I error (false positive rate).

Conservative Test A statistical test for which the Type I error rate (false positive rate) is actually less than that specified for the test. For a conservative test there will be a greater tendency to accept the null hypothesis when it is not true than for a non-conservative test.

Distribution The frequencies (either relative or absolute) with which measurements in a data set fall within specified classes. A graphical display of a distribution is referred to as a *histogram.*

Estimate Any numerical quantity computed from a sample of data. For example, a sample mean is an estimate of the corresponding population mean.

False Positive Rate The probability of mistakenly concluding that the sample area is clean when it is dirty. It is the probability of making a Type I error.

False Negative Rate The probability of mistakenly concluding that the sample area is dirty when it is clean. It is the probability of making a Type II error.

Geostatistics A methodology for the analysis of spatially correlated data. The characteristic feature is the use of variograms or related techniques to quantify and model the spatial correlation structure. Also includes the various techniques such as kriging, which utilize spatial correlation models.

Histogram A graphical display of a frequency distribution.

Hot Spot Localized elliptical areas with concentrations in excess of the cleanup standard, either a volume defined by the projection of the surface area through the soil zone that will be sampled or a discrete horizon within the soil zone that will be sampled.

APPENDIX D: GLOSSARY

Hypothesis An assumption about a property or characteristic of a population under study. The goal of statistical inference is to decide which of two complementary hypotheses is likely to be true. In the context of this guidance document, the *null* hypothesis is that the sample area is "dirty" and the *alternative* hypothesis is that the sample area is "clean."

Inference The process of generalizing (extrapolating) results from a sample to a larger population.

Judgment sample A sample of data selected according to non-probabilistic methods.

Kriging A weighted-moving-average interpolation method where the set of weights assigned to samples minimizes the estimation variance, which is computed as a function of the variogram model and locations of the samples relative to each other, and to the point or block being estimated. This technique is used to model the contours of contamination levels at a waste site (see Chapter 10).

Less-Than-Detection-Limit A concentration value that is below the detection limit. It is generally recommended that these values be included in the analysis as values at the detection limit.

Lognormal Distribution A family of positive-valued, skewed distributions commonly used in environmental work. See Gilbert (1987, p.152) for a detailed discussion of lognormal distributions.

Mean The arithmetic average of a set of data values. Specifically, the mean of a data set, $x_1, x_2, \ldots, x_n$, is defined by $\bar{x} = \sum_{i=1}^{n} x_i/n$.

Median The "middle" value of a set of data, after the values have been arranged in ascending order. If the number of data points is even, the median is defined to be the average of the two middle values.

Nonparametric Test A test based on relatively few assumptions about the underlying process generating the data. In particular, few assumptions are made about the exact form of the underlying probability distribution. As a consequence, nonparametric tests are valid for a fairly broad class of distributions.

Normal Distribution A family of "bell-shaped" distributions described by the mean and variance, μ and σ^2. Refer to a statistical text [(e.g., Sokal and Rohlf (1973)] for a formal definition. The *standard normal distribution* has $\mu = 0$ and $\sigma^2 = 1$.

Null Hypothesis See *hypothesis*.

Ordinary Kriging A variety of kriging which assumes that local means are not necessarily closely related to the population mean, and which therefore uses only the samples in the local neighborhood for the estimate. Ordinary kriging is the most commonly used method for environmental situations.

Outlier A measurement that is extremely large or small relative to the rest of the data gathered and that is suspected of misrepresenting the true concentration at the sample location.

Parameter A statistical property or characteristic of a population of values. Statistical quantities such as means, standard deviations, percentiles, etc. are parameters if they refer to a population of values, rather than to a sample of values.

Parametric Test A test based on relatively strong assumptions about the underlying process generating the data. For example, most parametric tests assume that the underlying data are normally distributed. As a consequence, parametric tests are not valid unless the underlying assumptions are met. See *robust test*.

Percentile The specific value of a distribution that divides the set of measurements in such a way that P percent of the measurements fall below (or are equal to) this value, and 1-P percent of the measurements exceed this value. For specificity, a percentile is described by the value of P (expressed as a percentage). For example, the 95th percentile (P=0.95) is that value X such that 95 percent of the data have values less than X, and 5 percent have values exceeding X. By definition, the median is the 50th percentile.

Physical sample or **soil sample** A portion of material (such as a soil core, scoop, etc.) gathered at the waste site on which measurements are to be made. This may also be called a **soil unit**. A soil sample may be mixed, subsampled, or otherwise handled to obtain the sample of soil that is sent for laboratory analysis.

APPENDIX D: GLOSSARY

Point Estimate See *estimate*.

Population The totality of soil units at a waste site for which inferences regarding attainment of cleanup standards are to be made.

Power The probability that a statistical test will result in rejecting the null hypothesis when the null hypothesis is false. Power = 1 - β, where β is the Type II error rate associated with the test. The term "power function" is more accurate because it reflects the fact that power is a function of a particular value of the parameter of interest under the alternative hypothesis.

Precision See *standard error*.

Proportion The number of soil units in a set of soil units that have a specified characteristic, divided by the total number of soil units in the set. This may also be expressed as a proportion of area or proportion of volume that has a specified characteristic.

Random Sample A sample of soil units selected using the simple random sampling procedures described in Chapter 5.

Range The difference between the maximum and minimum values of measurements in a data set.

Robust Test A statistical test that is approximately valid under a wide range of conditions.

Sample Any collection of soil samples taken from a waste site.

Sample Area The specific area within a waste site for which a separate decision on attainment is to be reached.

Sample Design The procedures used to select the sample of soil units.

Sample Size The number of lab samples (i.e., the size of the statistical sample). Thus, a sample of size 10 consists of the measurements taken on 10 lab samples.

Sequential Test A statistical test in which the decision to accept or reject the null hypothesis is made in a sequential fashion. A sequential test for proportions is described in Chapter 8 of this guidance document.

Semi-variogram Identical to the term "variogram." There is disagreement in the geostatistical literature as to which term should be used.

Significance Level The probability of a Type I error associated with a statistical test. In the context of the statistical tests presented in this document, it is the probability that the sample area is declared to be clean when it is dirty. The significance level is often denoted by the symbol α (Greek letter alpha).

Size of the physical sample This term refers to the dimensions of a physical sample or soil unit.

Skewed Distribution Any nonsymmetric distribution.

Soil Sample See *physical sample*.

Standard Deviation A measure of dispersion of a set of data. Specifically, given a set of measurements, $x_1, x_2, \ldots, x_n$, the standard deviation is defined to be the quantity, $s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$, where $\bar{x}$ is the sample mean.

Standard Error A measure of the variability (or precision) of a sample estimate. Standard errors are often used to construct confidence intervals.

Statistical Sample A collection of chemical concentration measurements reported by the lab for one or more lab samples.

Statistical Test A formal statistical procedure and decision rule for deciding whether a sample area attains the specified cleanup standard.

Stratified Sample A sample comprised of a number of separate samples from different strata.

APPENDIX D: GLOSSARY

Stratum A subset of a sample area within which a random or systematic sample is selected. The primary purpose of creating strata for sampling is to improve the precision of the sample design.

Symmetric Distribution A distribution of measurements for which the two sides of its overall shape are mirror images of each other about a center line.

Systematic Sample A "grid" sample with a random start position.

Tolerance Interval A confidence interval around a percentile of a distribution of concentrations.

Type I Error The error made when the sample area is declared to be clean when it is contaminated. This is also referred to as a *false positive*.

Type II Error The error made when the sample area is declared to be dirty when it is clean. This is also referred to as a *false negative*.

Variance The square of the standard deviation.

Variogram A plot of the variance (one-half the mean squared difference) of paired sample measurements as a function of the distance (and optionally of the direction) between samples. Typically, all possible sample pairs are examined, distance and direction. Variograms provide a means of quantifying the commonly observed relationship that samples close together will tend to have more similar values than samples far apart.

Waste Site The entire area being investigated for contamination.

Z Value Percentage point of a standard normal distribution.

INDEX OF KEY WORDS

alpha, 2-5
analysis plan, 4-6, 9-8
ARAR (applicable and relevant or appropriate requirements), 1-1, 1-7
beta, 2-5
Cs (cleanup standard), 3-5, 6-1
compositing methods, 5-15, 6-26
data quality objectives, 3-11
degrees of freedom, 6-3, 6-24, 6-28
exact test for proportions, 7-9, 7-10
false negative, 2-5, 2-7
false positive, 2-5, 2-7, 3-11
geostatistical methods, 10-1
grids, 5-5
homogenization, 10-6
hot spots, 9-1
kriging, 10-2
  indicator kriging, 10-9, 10-10
  probability kriging, 10-9, 10-10
mean concentration, 2-8, 2-9, 3-3, 3-10, 6-1, 6-16
median, 3-6
null hypothesis, 2-3
outliers, 2-16
parametric procedures, 6-1
proportions or percentiles, 2-9, 2-15, 7-1
quality assurance/quality control, 5-18
SARA (Superfund Amendments and Reauthorization Act of 1986), 1-1
sample
  judgment sample, 2-18, 4-2
  random sample, 2-18, 4-2, 4-4, 5-3
  stratified sample, 2-18, 4-4
  systematic sample, 2-18, 4-2
sample area, 3-1
sample size determination, 2-14, 6-7, 6-8
  for random samples (mean), 6-7
  for stratified random samples (mean), 6-13,
  for systematic samples (mean), 6-20
  for simple random samples (proportions/percentiles), 7-5
  for stratified samples (proportions/percentiles) 7-13, 7-17
  for a normal or lognormal population using tolerance intervals (proportions/percentiles), 7-21
sampling, 4-1
  random sampling, 4-1, 4-2, 7-5
  sequential sampling, 4-2, 4-6, 8-1
  stratified sampling, 4-1, 4-4, 6-12, 7-12
  systematic sampling, 4-1, 4-2, 6-21, 6-22, 6-23
sampling location, 5-1
  random samples, 5-3
  stratified samples, 5-13
  systematic samples, 5-5
sampling plan, 4-1
serpentine pattern, 6-21, 6-25, 6-26, 6-27
simple exceedance rule method, 7-11
soil unit, 2-18
spatial characterization, 1-4
standard deviation, 6-2, 6-4
standard error
  estimation of, 6-17, 6-18, 6-21, 6-23, 6-25, 7-7, 7-18
standards, 1-6
  background-based, 1-7
  cleanup, 1-6, 3-5
  risk-based, 1-8, 2-11, 2-12
  technology-based, 1-7
strata, 4-4
subsampling, 5-14
Superfund remediation, 1-6
upper percentile, 3-6
vadose zone, 1-4

*U.S. GOVERNMENT PRINTING OFFICE:1989-517-003:04871